Full Length Article

Gains and losses of exclusivity in grocery retailing

Katrijn Gielen, Els Gijsbrechts, Marnik G. Dekimpe

Abstract

Conventional wisdom dictates that convenience goods should be distributed as intensively as possible. Still, exclusivity arrangements are rapidly gaining way in grocery retailing. We discuss the possible performance outcomes of exclusivity deals, and propose a unified framework (i) to quantify the gains and losses of such arrangements in consumer-packaged-goods markets, and (ii) to decompose the total monetary gains and losses into a variety of sources. Our framework considers both the manufacturer and the retailer granted the exclusivity right, and accounts for the fact that both dyad parties may be active in multiple, inter-related, categories.

We illustrate the proposed approach in the context of Unilever’s decision to limit the distribution of five of its (sub-)brands to a single retailer in the Dutch market, and derive the sales and profit implications for the parties involved. In all instances, we find that intensive distribution generates higher sales for the manufacturer than the exclusive arrangement, while such an arrangement is typically appealing for the retailer granted exclusivity. To come to a win-win situation, a variety of compensatory arrangements are considered. While manufacturers should not try to have leading incumbent products de-listed, we show how extra feature support for the exclusive brand may make an otherwise harmful arrangement beneficial for the manufacturer. In addition, we show how, in a number of instances, the total dyad profitability increases under exclusive arrangements, which offers rooms for margin re-negotiations.

1. Introduction

A key development in grocery retailing is the increase in multi-store patronage, making strict store loyalty an exception. In response, retailers try to differentiate themselves. As most grocery stores carry the same categories, differentiation in assortment composition depends almost entirely on variation within the respective categories (Briesch, Chintagunta, & Fox, 2009). Traditionally, retailers have tried to achieve within-category differentiation through their private labels. However, recent evidence suggests that there are limits to how far retailers can push their private-label program (Ailawadi, Lehmann, Neslin, Pauwels, & Steenkamp, 2008). As an alternative differentiation strategy, retailers can try to obtain exclusive distribution rights from key national brands.

As such, Tesco obtained exclusive rights to distribute Procter & Gamble’s hair care range Physique (Hayward, 2002) and Unilever’s Vaseline Cocoa Butter range in the UK (Kiss & Makeup, 2008). In the US, PepsiCo launched its new energy drink Fuelosophy through Whole Foods (Thompson, 2006), while Estée Lauder created four new brands (American Beauty, Flirt, Good Skin and Grassroots) to be sold exclusively at the mid-priced department-store chain Kohl’s (Kumar, 2007).

Retail exclusivity is already very popular among toy manufacturers and with apparel brands. Even though exclusivity used to be rare in grocery markets (Sudhir & Rao, 2006, p. 153), industry observers (PricewaterhouseCoopers/TNS Retail Forward, 2007) expect this practice to increase dramatically in that sector as well. Procter & Gamble, for example, recently announced that it wants to expand its number of exclusive brands and variants sold exclusively at select retailer partner stores (Failla, 2010). In a 2007 GfK survey for the UK Competition Commission among over 450 grocery-product suppliers, 35% admitted that they had been asked to enter into an exclusivity agreement with one of their retail customers. Two-thirds of those requests, which did not involve private-label production but dealt with the exclusive distribution of branded products, came from one of the four largest supermarkets. Overall, 19% of all suppliers actually entered into at least one such exclusivity agreement.

This practice goes against conventional wisdom (e.g., Coughlan, Anderson, Stern, & El-Ansary, 2001) that convenience goods “should be distributed as intensively as possible” (p. 288, italics added), as buyers will not expend much effort to purchase them. Exclusivity, by definition, limits the product’s exposure to one retailer. Unlike private labels, however, the brand remains owned, branded, and marketed by the manufacturer, as is also clearly communicated to the consumer. As

Keywords:
Retailing
Distribution intensity
Exclusivity
Consumer packaged goods
Marketing accountability

Article history:
First received in August 10, 2012 and was under review for 10 months
Available online 24 April 2014

Area Editor: Kusum L. Ailawadi

© 2014 Elsevier B.V. All rights reserved.

http://dx.doi.org/10.1016/j.ijresmar.2014.03.003
0167-8116/© 2014 Elsevier B.V. All rights reserved.
such, it retains the identity of the manufacturer. The practice also differs from the use of branded variants, studied in Bergen, Dutta, and Shugan (1996), which basically describes SKU customization where different variants of the same line are offered through different retailers. In our setting, only one retailer is allowed to carry an entire (sub-)branded product line.

Even though exclusive agreements in grocery retailing are becoming more prevalent, little is known on how to quantify the gains and losses, what drives the size of these, or the division between the different parties involved. In this paper, we develop a formal approach to assess the accountability of exclusive distribution. This will help national-brand managers decide which retailer requests for exclusivity to maintain/accept, under what conditions, and at what opportunity cost. We also explore some options that managers can pursue to amplify their gains (attenuate their losses). This is a challenging task.

First, multiple channel parties are involved, varying in terms of channel power, and both vertical and horizontal conduct between these parties will shape the overall outcome. A priori, one expects the retailer to gain from exclusive arrangements, while manufacturers face lost sales opportunities when distribution is limited. This is consistent with O'Brien and Shaffer’s (1997) argument that retailers prefer an equilibrium with exclusivity, while manufacturers prefer an equilibrium without exclusivity. However, this argument ignores that the dyad’s overall profitability can increase. Because of this, renegotiations (e.g., in terms of trade support and/or margin division) may result in win-win situations.

Second, even if the exclusivity agreement pertains to a specific brand and/or product category, its consequences may extend well beyond the boundaries of that brand and category. Retailers carry multiple categories, which may exhibit complementary or substitution relationships. As such, they should not only strive to maximize the performance of the category in which exclusivity rights are gained. Instead, they are interested in overall store (chain) performance, i.e., across all categories (Chen, Hess, Wilcox, & Zhang, 1999). Likewise, many manufacturers are active in multiple categories. They, too, may want to also consider the cross-effects with other brands in their portfolio.

In this paper, we develop an empirical framework that captures the intricacies of the grocery retailing’s institutional context, and that allows to (i) quantify the total monetary implications of exclusivity to both manufacturers and retailers, (ii) decompose this total effect into its key components, and (iii) assess the performance ramifications of alternative compensation schemes from one dyad partner to the other.

2. Conceptual background

Prior research has looked at the underlying reasons why retailers typically favor exclusivity. Having a unique product in one’s assortment can lead to increased store traffic by capturing market share from retailers not carrying the product, and generate a positive spill-over effect when consumers also buy other products while in the store. Assortment composition has been identified as a key determinant of store choice and retail success (see, e.g., Grewal, Krishnan, Levy, & Munger, 2010; Mantrala et al., 2009), and exclusive distribution arrangements can be used to draw additional customers to the store. Once in the store, other categories may be affected as well, as basket-shopping customers may want to avoid extending their search to other stores (Cachon & Kök, 2007). Customers then not only buy in the focal category, but also complement these purchases with products from other categories (Seetharaman et al., 2005), thereby positively impacting the focal retailer’s performance. Moreover, given that the product is not available at other stores, intra-brand competition (between different retailers carrying the brand) that could drive the price down is avoided (Andritsos & Tang, 2010; Cheng, 2008), allowing the retailer to extract monopoly rents from the market (Matouschek & Ramezanna, 2007). Given these observations, exclusivity is widely assumed to lead to higher retailer revenues and profits, and the extent of these gains has been shown to increase with the level of competition, and to decrease with the level of differentiation (see, e.g., Chen, Iyer, & Padmanabhan, 2002; Soberman, 2009).

Manufacturers, in turn, tend to be more ambivalent. On the one hand, exclusivity can offer the manufacturer more control opportunities (Frazier & Lassar, 1996) and/or serve as a way to signal his commitment to the selected channel partner, which could elicit subsequent retailer support (Fein & Anderson, 1997). On the other hand, there is the concern that sales opportunities will be lost when distribution is restricted (Andritsos & Tang, 2010). The latter is especially relevant when the target market has a limited willingness to engage in search activities to obtain the firm’s products (Heide, 1994). Because of this, conventional wisdom tends to dictate that grocery products should be distributed as intensively as possible (Coughlan et al., 2001), if only because of the well-established convex relationship between the extent of retail distribution and volume market share (see, for example, Reibstein & Farris, 1995). Moreover, limiting the extent of distribution may reduce cross-distributor word-of-mouth communication, which could hurt the acceptance and diffusion of the new product (Peres & Van den Bulte, 2014).

By accepting an exclusive arrangement, the manufacturer essentially trades off market coverage with investments made in the partnership by the retailer. For exclusive arrangements to emerge/sustain, however, it is important that both parties find the arrangement mutually profitable (Subramanian, Raju, & Zhang, 2013). To that extent, compensatory measures may be called for in which one party (typically the retailer) shares part of its revenue gain with the other party (the manufacturer).

Many studies on exclusivity (see, e.g., Engene & Parry, 1995 or Trivedi, 1998 for a review) have developed analytical (game-theoretic) models to explore, in a stylized setting, under what conditions exclusive deals can become an equilibrium strategy for manufacturers and retailers. In this respect, prior research has considered exclusivity contracts for advertising (Dukes & Gal-Or, 2003), for referral infomediaries (Chen et al., 2002), and for handset arrangements in the wireless industry (Subramanian et al., 2013), among others. Cai, Dai, and Zhou (2012), in turn, considered the use of exclusive channels and revenue sharing/bargaining solutions in complementary-goods markets. However, as already observed by Fein and Anderson (1997, p. 22), key results obtained in this research tradition are (i) sensitive to the specific institutional context (and associated modeling assumptions), and (ii) difficult to test empirically. For mathematical tractability, these studies tend to consider a limited (often two) number of retailers and/or manufacturers. However, while a focus on a two-player setting can capture the key features of some markets (as the Telecom industry in Cai et al., 2012), they are not well suited for the multi-supplier, multi-retailer, and multi-category setting typically seen in CPG markets. This calls, as Cai et al. (p. 182) point out, for a more empirical treatment, as it is very difficult to capture this high dimensionality in a game-theoretic model, and/or to derive useful analytical insights for such a complex institutional context.

In response to this call, our modeling framework will offer three key contributions. First, we quantify the total monetary impact for both parties entering an exclusive arrangement, explicitly accounting for the following features of the grocery market: (i) multiple retailers are competing with one another, and carry both their own store brand (private labels) and national brands coming from a variety of manufacturers; (ii) both retailers and manufacturers are active in multiple categories, which may exhibit not just complementary but also substitution relationships, and (iii) retailers differ along multiple dimensions, such as the price of the products sold, their promotional intensity and (especially relevant in our context) the assortment offered.6 Exclusive arrangements alter the assortment composition in a given category with a given retailer. This may have performance implications not

---

6 Analytical models, in contrast, tend to focus on a single instrument (typically price).
only within that category/retailer, but also in other categories and/or with competing retailers. The latter effects should be taken into account to come to a full ramification of an exclusive deal’s consequences. Such insights should be useful for both parties to evaluate existing exclusivity arrangements, and – once clear patterns emerge across several of such cases – allow to evaluate proposed deals (typically by the retailer to the manufacturer) upfront.

Second, we decompose this total monetary impact into its constituent parts. For the retailer, gains/losses can be realized not only in the focal category, but also in a variety of other categories. For the manufacturer, the situation is even more intricate, as there may be implications not only with the focal retailer, but also with other retailers selling some of its other products, again across multiple categories. A decomposition into these various sources will offer additional insights that can be useful to manufacturers when deciding whether or not to accept a proposed deal, and/or help in anticipating potential competitive reactions.

Third, in case of opposing effects for both parties in the dyad, we consider to what extent there is room to negotiate a win–win situation. Three types of compensatory benefits will be considered that could make the exclusive deal more attractive to the manufacturer: (i) competitive delisting, (ii) altered feature support, and (iii) margin re-negotiations.

Anecdotal evidence suggests that manufacturers may try to take advantage of privileged relations with a retailer (e.g., by being the category captain) to drive out competing manufacturers at the point-of-sales (Glazer, Henry, & Jacobson, 2004). In return for exclusivity, the focal manufacturer may request the same. Indeed, he may expect that exclusive introductions which drive competing brands out of the retailer’s assortment (pending on whether the exclusive offering expanded the assortment, or whether it replaced competing brands (Mason, 1990)), the terms of this support may strongly affect the appeal of the exclusive deal more attractive to the manufacturer: (i) competitive advantage in the focal category and other categories. For the manufacturer, the situation is even more intricate, as there may be implications not only within that category/retailer, but also in other categories and/or brands (Broniaczyk & Hoyer, 2010).

Alternatively, rather than trying to weaken its competitors, the manufacturer could attempt to negotiate better terms for his own items as part of the exclusivity deal. For example, he can aim for enhanced support in the store flyer (preferably at the expense of the retailer). Given the huge budgets typically involved in trade support (Narasimhan, 2009), the terms of this support may strongly affect the appeal of the exclusive deal for both parties involved.

Finally, revenues can be transferred through a margin shift so that each unit sold becomes more profitable to the manufacturer. A key consideration in a manufacturer’s decision to engage in an exclusive arrangement is the sales that will be foregone when limiting the distribution intensity (Frazier & Lassar, 1996). However, the profit implications of an exclusivity arrangement are determined not only by the size of those foregone sales, but also by how both parties divide the total margin in the vertical dyad, as also reflected in several of the analytical (economic) models referred to before (e.g. Andritsos & Tang, 2010; Trivedi, 1998).

While revenue sharing and/or compensatory benefits have been discussed analytically as a way to make exclusive deals Pareto efficient for the dyad (see, e.g., Cai et al., 2012), we are the first to empirically consider all three mechanisms in a grocery-retail setting. Not only will we assess, across five actual exclusivity arrangements, (i) whether there was any room for revenue sharing between the dyads’ parties, but we will also gauge (ii) whether there is any empirical evidence for the actual implementation of the respective mechanisms, and if so, (iii) quantify the resulting performance implications and assess whether they can imply a win–win. If so, some of the manufacturers’ initial concerns when asked to join an exclusivity arrangement may become alleviated.

3. Methodology

To assess the gains and losses of the exclusive introduction of a manufacturer brand line for both the manufacturer (m) and the exclusive retailer (r) in a dyad, we contrast the sales resulting from (i) an exclusive distribution arrangement (labeled E), and (ii) an intensive distribution scenario through all leading retailers (I). In both settings, sales gains are evaluated in comparison to a pre-introduction ‘null’ scenario. To adequately evaluate the sales gains of exclusive arrangements, we each time take into account that the introduction may affect the focal retailer and manufacturer sales not only in the focal category (c), but also in other categories in which the retailer and manufacturer are operating, and that manufacturer sales can be affected at other retailers not included in the exclusive arrangement. Fig. 1 gives an overview of the different components and their interrelationships.

Formally, retailer sales changes can be written as the sum of the sales effect in the focal category and other categories (cf. Fig. 1, Panel B1):

\[
\text{Gain}_{cr}^E = \left( \frac{e_{rE} - e_{rI}}{C_{16}C_{17}} \right) + \sum_{c \neq c} \left( \frac{e_{cE} - e_{cI}}{S_{1}C_{17}} \right)
\]

Note that the resulting differences can be negative, in which case the retailer incurs an exclusivity loss.

In a similar vein, manufacturer sales gains can be written (cf. Fig. 1, Panel B2) as the sum of the sales gains in the focal category, the foregone sales the manufacturer incurs when not selling the brand at retailers other than r, and potential cross-category effects:

\[
\text{Gain}_{cr}^M = \left( \frac{\text{ManShare}_{crE}^r + S_{1}C_{17} - \text{ManShare}_{crI}^r \times S_{1}C_{17}}{S_{1}C_{17}} \right) + \sum_{c \neq c} \left( \frac{\text{ManShare}_{cE}^m \times S_{1}C_{17} - \text{ManShare}_{cI}^m \times S_{1}C_{17}}{S_{1}C_{17}} \right)
\]

\[
+ \sum_{c \neq c} \sum_{r \neq r} \left( \frac{\text{ManShare}_{cE}^m \times S_{1}C_{17} - \text{ManShare}_{cI}^m \times S_{1}C_{17}}{S_{1}C_{17}} \right)
\]

Three differences relative to the retailer gains have to be pointed out. First, for the manufacturer only the portion of the category attributed to its own brands is relevant. We therefore have to factor in what share (ManShare) of each retailer’s category sales (S_{1}) is captured by the manufacturer. Second, not only sales at the focal retailer r, but also sales at all retailers (R^{m}) that carry the focal manufacturer’s brands have to be considered. Third, only categories in which the manufacturer operates (C_{r}^{m}) are included.

To quantify Gain_{cr}^{E} and Gain_{cr}^{M}, we proceed as follows:

(a) Distribution intensity operationalization. We study the implications of differential distribution intensity by operationalizing the assortment in each of the relevant stores through a number of key elements related to the introduction (Fig. 1, Panel A): number of brands, number of sub-brands per brand, and number of SKUs per brand. Indeed, exclusive vs. intensive distribution is reflected in different values for these assortment variables across the different retailers, which, in turn, can cause changes in performance (Fig. 1, Panels B1–B2). Exploiting the observed variation in those assortment variables across categories, retail

4 Each of the items in Eq. (1) can also be written as \((S_{E} - S_{I}) - (S_{E} - S_{I})\), which emphasizes the differential sales level relative to the no-introduction scenario. A similar argument applies to the difference terms in Eq. (2) below.
Performance Implications for the Dyad

(a) Sales estimation. To obtain reliable estimates of retailer–category sales under the different scenarios, several challenges must be met. First, introductions in one category at one retailer may affect the relative appeal of other categories and retailers, and our model should properly capture these interdependencies. Second, since categories can be substitutes as well as complements, we need a model that accommodates both types of relationships. Third, exclusive arrangements affect not only the retailer’s sales composition, but also his share of the total market—a key metric for retailer performance. Conversely, the manufacturer does not necessarily have an interest in total retailer–category sales, but rather in the portion captured by his own brands. Hence, netting out the exclusivity gains for each channel partner calls for separate model components to capture their specific stakes. With this in mind, we decompose the retailer–category sales $S_{cr}$ into three components: the category development index at the retailer, the retailer’s sales share, and the category sales in the total market: $S_{cr} = CDL_{cr} \times (S_{cr}^{f}) \times S_r$.

Similar in spirit to writing brand sales as the product of brand market share and category sales (see, e.g., Leeflang, Wittink, Wedel, & Naert, 2000, pp. 102–103), this decomposition allows to address the above challenges as outlined in the next section. To obtain the manufacturer’s category sales, we will multiply $S_{cr}$ with an additional manufacturer–share component: $\text{ManShare}^7$. Using historical data, the different subcomponents will be related to the marketing-mix instruments that entail exclusive introductions. More detail on the models used to capture the different subcomponents is provided in Sections 3.1 and 3.2.

(b) Sales estimation. To obtain reliable estimates of retailer–category sales under the different scenarios, several challenges must be met. First, introductions in one category at one retailer may affect the relative appeal of other categories and retailers, and our model should properly capture these interdependencies.

(c) Distribution arrangement simulations. We use the estimated parameters to derive conditional forecasts for (i) a pre-introduction setting, (ii) an exclusive distribution scenario, where the product line is exclusively distributed through one retailer, and (iii) an intensive distribution scenario, where the product line is distributed through all retailers. Similar “restricted policy simulations” have been used in Pauwels (2004), Horváth, Leeflang, Wierenga, and Wittink (2005) and Osinga, Leeflang, Srinivasan, and Wieringa (2011) among others. To derive sales predictions for the pre-introduction scenario, we fix all marketing-mix variables to their level preceding the introduction, and use our subcomponent models to predict the corresponding sales. As such, we isolate the product-line-availability effect from other market changes. We then compare the focal retailer’s performance in this null case, ceteris paribus, with the results predicted for the first year under, respectively, exclusive ($S_{cr}^{f} - S_{cr}^{E}$) and intensive ($S_{cr}^{f} - S_{cr}^{I}$) distribution. A similar approach is used for the manufacturer. By comparing predicted metrics, we avoid a confound with model forecast errors (see also Szymanowski & Gijssbrechts, 2012).

(d) Allowing for compensatory benefits. Finally, we simulate the performance implications for both dyad partners of three sets of compensatory benefits: (i) competitive de-listings, (ii) additional feature support, and (iii) a re-allocation of the dyad’s margin (Panel C in Fig. 1), and assess to what extent they can render the exclusivity deal a win–win proposition for both the manufacturer and the retailer.

---

7 As an alternative, one might consider the difference-in-difference approach used in Ailawadi et al. (2010). However, given that an exclusive deal is typically offered in all stores of a (nationwide) chain, no control group (region) is available—thereby precluding that approach. Such a lack of control group was also an issue in other studies of changes in consumer behavior following a natural experiment with only a before and after condition. Examples include Procter & Gamble’s switch to value pricing (Ailawadi, Lehmann, & Neslin, 2001) and the occurrence of the Dutch price war (van Heerde, Dekimpe, Potapov, Gijssbrechts, & Pauwels, 2008). However, the methods used in those studies would not be suitable in our setting either: we are interested not only in comparing the situation prior to the exclusive introduction (the pre-condition) with the situation following that introduction (the post-condition), but also (and even more so) in comparing the latter with the (non-observed) situation of intensive distribution.
Below, we provide more technical detail on the models used to obtain the conditional forecasts.

3.1. Retailer sales

Retailer–category sales, as indicated before, are decomposed into three subcomponents (i) the category development index at the retailer, (ii) the retailer’s sales share, and (iii) the category sales in the total market. Whereas the latter component of retailer–category sales, i.e. category sales in the total market (i.e. country-wide, across all retailers), is not likely to be affected by exclusive introductions, as was confirmed by structural-break tests, the other components will change following the introduction. Below, we discuss how we relate both components to the relevant marketing-mix instruments.

3.1.1. Category development index

The first retailer–category sales component, CDI, is defined as the ratio of a retailer’s share in a given category to his share of the total market

\[ \text{CDI}_{r,t} = \frac{\text{Sales}_r/c_t}{\text{Sales}_r/c_t} \]

where \( S \) represents sales value, \( t \) represents time, and \( c = 1, \ldots, C, t = 1, \ldots, T \) refer to the category and the retailer, respectively (the dot (‘.’) refers to a summation).

The CDI component allows us to efficiently capture cross-effects across both categories and retailers. Traditional metrics (profits, sales, market share) focus on one dimension (either the retailer or category dimension). In contrast, the CDI metric accounts for both simultaneously (Dhar, Hoch, & Kumar, 2001). For example, some retailers may have more expertise (a stronger position) in high-quality produce and fresh meat (but less in cosmetics), while other retailers may put more emphasis on the latter (see also Dragnska & Klapper, 2007). The CDI metric has not only found its way into the academic literature, but it is also a commonly-used metric in the trade. A value of one indicates that the retailer realizes his fair share in the category, while values above (below) one reflect superior (inferior) performance.

We relate the respective CDIs to the characteristics for each retailer and category that capture assortment changes, while controlling for price and promotion actions. We denote the \( M \) marketing-mix instruments by \( \chi_{r,m,k}^{\text{m}} \) (\( k: 1, \ldots, K \)). Changes in assortment variables like the number of SKUs offered by a brand at a given retailer allow us to factor in the impact of exclusive distribution decisions. The impact of these variables can differ between private labels and national brands, and between different national-brand manufacturers (as denoted by superscript \( m, m = 1, \ldots, M \)). Instead of considering absolute levels of these variables, we express them in ‘DI’ or development-index format as well. If \( \chi_{r,m,k}^{\text{m}} \) is the absolute level of variable \( k \), for manufacturer \( m \), retailer \( r \) and category \( c \) in period \( t \), the corresponding DI-variable is defined as:

\[ \chi_{r,m,k}^{\text{DI}} = \frac{\chi_{r,m,k}^{\text{m}}}{\chi_{c,k}}. \]

This specification has intuitive appeal: like the CDI performance measure, it offers an indication of the relative level of marketing inputs allocated to a specific category and retailer.7

In line with earlier work (Dhar et al., 2001), we use a multiplicative function to link these marketing inputs to the CDIs. Even though such a function offers an efficient way to capture substitution between category–retailer CDIs, it does not allow for complementarities between categories. We therefore follow Carpenter, Cooper, Hanssens, and Midgley (1988) and Vanden Abeele, Gijsbrechts, and Vanhuele (1990), and allow selected cross-category effects to deviate from the overall pattern, and to even reflect complementary relationships:

\[
\text{CDI}_{r,t} = \alpha_c \alpha_r \prod_{k} \prod_{\text{m}} \left( \frac{\chi_{r,m,k}^{\text{DI}}}{\chi_{c,k}} \right)^{\beta_{c,k}^m} \text{RetShare}_{r,t}^{\text{m}} \text{e}^{\Psi_{r,t}^m} \]

where \( \text{RetShare}_{r,t}^{\text{m}} \) is a normally-distributed error term with mean zero and variance \( \sigma_{\text{RetShare}}^2 \). Retailer differences are captured through retailer-specific fixed effects, \( \alpha_r \) (for identification, we denote one retailer as the ‘reference retailer’, for which \( \alpha_r = 1 \)). Given the much larger number of categories (which renders the fixed-effects model less feasible), unobserved heterogeneity between categories is accommodated by means of a random-effects specification. Specifically, we use normal mixing distributions for the scale parameter: \( \alpha_c \sim N(\tau, \sigma_{\alpha_c}^2) \) – with \( \sigma_{\alpha_c}^2 \) as cross-category variance – as well as for the slope parameters related to the marketing-mix instruments:

\[ \beta_{c,k}^m \sim N(\beta_{c,k}^0, \sigma_{\beta_{c,k}}^2) \]

\( \text{Cl}_k \) is the index set of categories of which the interdependency with category \( c \) deviates from the ‘base’ structure; and \( \Psi_{r,t}^m \) is the parameter reflecting this deviation. For categories not in the index set, it equals zero. Categories with \( \Psi_{r,t}^m \) greater (smaller) than zero, exhibit less pronounced (stronger) substitution relationships with category \( c \). As advocated by Carpenter et al. (1988, p. 399), we use a three-step procedure to identify the index set and obtain the parameters: we (i) first estimate the model without cross-effects, then (ii) for every category, link the residuals of that model to the other-category predicted attractions (calculated using the parameters of the first step), and then (iii) re-estimate after retaining only the significant cross-effects (p < .05) in the second step.

3.1.2. Retailer sales share

Marketing-mix changes can affect not only the relative appeal of different categories within a retailer, but also the overall position of that retailer. We specify the second component of retailer–category sales, the retailer’s sales share, \( \text{RetShare}_{r,t} \), in a given period, \( (S_r/S_c) \), as a function of his relative marketing efforts. Consistent with our CDI specification, we use a multiplicative model to link an instrument’s share \( X_{r,\text{Share}} \) to the retailer:

\[ \text{RetShare}_{r,t} = \gamma_r \prod_{\text{m}} \prod_{k} \left( \frac{X_{r,m,k}^{\text{Share}}}{\chi_{r,m,k}^{\text{m}}} \right)^{\delta_{r,m,k}^{\text{Share}}} \text{e}^{\Psi_{r,t}^m} \]

where \( \text{RetShare}_{r,t} \) is a normally-distributed error with mean zero and variance \( \sigma_{\text{RetShare}}^2 \). \( \delta_{r,m,k}^{\text{Share}} \) is a fixed-effect parameter for retailer \( r \), and \( \psi_{r,t}^m \) are the marketing-mix slope parameters. We do not need category-specific parameters here, since the variables pertain to the retailer as a whole.8 This (multiplicative) retailer–share model forms a natural extension to the (multiplicative) CDI model, allowing marketing-mix efforts to not just alter the relative performance of different categories within the chain, but to also affect the chain’s overall competitive position in the market.

3.2. Manufacturer sales

Within each category and retailer, manufacturer sales can be obtained by multiplying the retailer–category sales (\( S_{r,m} \)) with the portion of those sales accounted for by the manufacturer’s brands, \( S_{r,m}^{\text{m}} \). To obtain the latter, we allow the manufacturer share to be influenced by its own marketing-mix activities (including assortment

Note that, unlike the scale parameter, the slope parameters do not vary across retail chains. We also estimated a model including heterogeneity in the slopes across retailers, but found the deviations to be insignificant.

Moreover, as with the CDI model, we did not find evidence of significant heterogeneity in the slopes across retailers.
characteristics) as well as those of competitors, within the category and store.\textsuperscript{10} We use an attraction specification to specify what portion of retailer–category sales accrues to the different manufacturers — including the focal manufacturer:\textsuperscript{11}

\[
\text{ManShare}_{r,m} = \frac{\prod_{k=1}^{M} e^{\theta_{r,m,k}}}{\sum_{m=1}^{M} \prod_{k=1}^{M} e^{\theta_{r,m,k}}} e^{\text{Retailer,\textunderscore m}}.
\]

where \(u_{\text{Retailer,\textunderscore m}}\) is a normally-distributed error with mean zero and variance \(\sigma_{\text{Retailer,\textunderscore m}}^2\). \(\theta_{r,m,k}\) and \(\sigma_{r,m,k}\) are the brand-interpret and marketing-mix slope parameters, respectively. This model not only satisfies sum and range constraints, but it also incorporates manufacturer cross-effects in a parsimonious way. As before, we accommodate category- and retailer heterogeneity in these brand-interpret-parameters, by allowing for retailer fixed-effects, and using a normal mixing distribution for the category-specific component: \(\xi_{r,m,k} \sim N(0, \sigma_{r,m,k}^2\text{cat})\) (with \(\sigma_{r,m,k}^2\text{cat}\) the cross-category variance). For identification, we denote one retailer as the ‘reference retailer’, for which \(\xi_{r,m,k}\) is set to zero. Moreover, like in the CDI model, we allow the marketing-mix parameters \(\xi_{r,m,k}\) to vary randomly across categories:\textsuperscript{12} \(\xi_{r,m,k} = N(\xi_{r,m,k}, \sigma_{r,m,k}^2)\).

In sum, exclusive arrangement leads to shifts in assortment variables, which cause changes in retailer category sales (composed of retailer–category CDIs and retailer sales shares), and manufacturer category sales (composed of retailer category sales and manufacturer market shares). Having obtained the parameters of the retailer–share, manufacturer-share and CDI model (including the category cross-effect parameters), we calculate the predicted impact of alternative distribution scenarios, as explained above.

4. Empirical setting

4.1. Data

We use data provided by GfK Renelux\textsuperscript{13} on household spending in 53 grocery categories in the Netherlands. These categories represent more than 95% of the total CPG market. We consider purchases across all national retail chains: four chains of conventional supermarkets and two hard discounters. Purchases at other smaller retail outlets are combined in an “other retailers” group. Similarly, all purchases not covered by the 53 key categories are grouped in a “rest category”. Data are available for 104 4-weekly periods, between February 2002 and April 2010.

The focal firm, Unilever (UL), operated in 35 categories (e.g. hair care and ice cream). In 18 categories (e.g. pet food and breakfast cereals), UL was not (yet) active in the Dutch market. We also consider these categories, as they are relevant from the retailer’s perspective. The four traditional supermarkets in our data set are Albert Heijn, the market leader (market share: 27.7%), which has a high-price and high-service orientation, Super de Boer and Plus (market shares: 4.8% and 4.7%), which both score relatively high on price and assortment, and C1000, the value-oriented player among the Dutch supermarkets (market share: 11.3%). Aldi and Lidl (market share: 11.8% and 9.2%) are typical hard discounters carrying mostly private labels at bottom prices.

Our data represent different distribution-intensity scenarios. While many brands and sub-brands are available at all supermarkets, others are found in only one (some of the) chain(s). Table 1 provides some descriptive statistics for the national brands in our sample. It shows that while the majority of these brands’ sub-lines (i.e. their lines in specific product categories) are intensively distributed, an important portion (11.35%) is distributed through only two retailers, while an equally large portion (11.80%) is uniquely available at only one chain (e.g., Albert Heijn carries 61 exclusive brand-lines, and Super de Boer another 32) — which reflects the growing popularity of exclusive arrangements in a grocery setting. Moreover, exclusive lines are found in almost all (43 out of 53) categories. Hence, our data cover quite some variation in terms of distribution intensity. As such, we will stay well within the range of our data when varying the distribution-intensity variables in our policy simulations, in line with the recommendation of van Heerde, Dekimpe, and Putis (2005) to reduce the potential impact of the Lucas critique.

In these simulations, we will focus on five cases in which this exclusive distribution results from a deliberate strategy on the part of the manufacturer (UL): distribution rights being exclusively granted to one retailer (i.e. the leading retailer, Albert Heijn), in spite of requests by other retailers to also carry the product. Interestingly, the exclusive deals involve one of the world’s largest grocery retailers (number nine in terms of 2012 global grocery sales), and one of the largest CPG manufacturers (number four in the 2012 global ranking) — both of which therefore bring quite some power to the negotiation table.

An overview of these cases is provided in Table 2. In all five cases, exclusivity was granted on a long-term basis. The cases vary widely in the timing of the exclusive introduction: the earliest arrangement dates back to September 2003, while others occurred more recently. The exclusive deals are found in different categories (bread, cheese, hot drinks, ice cream), with widely different retailer shares (ranging from 15 to 32%) as well as UL shares prior to the introduction (ranging from zero and 48%). They also vary in the nature and scope of the introduction, i.e. whether the exclusive product was a new brand or a new sub-brand in the category, and the number of introduced SKUs. In addition, different implications for the incumbents’ assortment with that retailer are observed, as discussed later.

4.2. Operationalization of the performance measures

4.2.1. CDI

To operationalize the CDIs, we use Euro sales at the retailer and category level as provided by GfK. By definition, CDIs in each 4-weekly period are ‘benchmarked’ against one, and higher values for some categories (retailers) are compensated by lower values for others (Dhar et al., 2001). The CDIs vary considerably across both retailers and categories. In the fish category, for example, the values range between .73 (Super de Boer) and 1.29 (Albert Heijn). The corresponding values for the cleaning detergents category are 1.12 and .93, respectively.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Distribution intensity for national brands’ sub-lines.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of brand-lines available at</td>
<td>Overall</td>
</tr>
<tr>
<td>One chain</td>
<td>190</td>
</tr>
<tr>
<td>Two chains</td>
<td>182</td>
</tr>
<tr>
<td>Three chains</td>
<td>163</td>
</tr>
<tr>
<td>All chains</td>
<td>1074</td>
</tr>
<tr>
<td>Total</td>
<td>1609</td>
</tr>
</tbody>
</table>

Note: The table covers all (837) national brands available at the four chains (i.e. that have at least one sub-line in each supermarket), and reports the distribution intensity of these brands’ 1609 sub-lines. For instance, 182 out of the 1699 lines are sold through only two chains, and Albert Heijn carries 95 of those 182 lines.
4.2.2. Retailer sales share

The retailer’s share is based on the retailer’s sales value (in monetary terms), summed across all categories, and divided by total sales value across all retailers.

4.2.3. Market share

Each manufacturer’s market share in a retailer–category combination is captured as the ratio of the value sales of all products of that manufacturer in a category at a specific retailer, relative to the total category value sales at that retailer. On average, UL’s shares in the categories in which it operates amount to 11.2%, 11.6%, 13.7%, and 13.6% at AH, C1000, Plus, and Super de Boer. This relative position varies considerably across categories, as can be seen in Table 2 when comparing UL’s share across the different product categories. UL’s presence at the hard discounters is much more modest, with shares near zero. Indeed, discounters sell almost exclusively PLs.

4.3. Quantifying the marketing-mix variables

As explanatory variables, we include the number of brands, the number of sub-brands and SKUs per brand, the regular unit price relative to other brands in the category, the number of feature promotions, and the average depth of price promotions. Information on these variables was provided by GfK. Each variable is measured for the private label, UL, and the top two competing manufacturers in the category (NB1 and NB2) — the remaining national brands being grouped into an ‘other’ category. Similar to Srinivasan, Pauwels, Hanssens, and Dekimpe (2004), a brand is said to experience a price promotion when its price in a given week is more than one standard deviation below the regular retail price level. Price-promotion depth is defined as the percentage difference between the promotional price and the regular price. By averaging this value across all promotional weeks in a given 4-week period, promotional depth in a retailer’s category is obtained. The amount of feature support is obtained from GfK. Table 3 provides summary statistics for all variables in the focal categories, and underscores that substantial variability exists, both cross-sectionally (across brands, categories and retailers) and over time.14

All marketing-mix variables enter the CDI model and the retailer’s share model after proper transformation, i.e. a DI transformation in the CDI model and a share transformation in the retailer share model, while absolute values within the category and retailer are used in the manufacturer attraction model.

All marketing-mix variables are instrumented using a set of variables that reflect evolutions in competition, demand and production costs, and/or allow for systematic differences between categories (in line with the suggestions of Srinivasan, Pauwels, & Nijs, 2008 and Luan & Sudhir, 2010, among others). As competition-based instruments we consider the values of the focal marketing variable (i) for the focal brand, within the same category, averaged across the competing retailers and (ii) within the same category and retailer, but across the competing brands.15 For both instruments, we use their lagged value (cf. Albers, 2012). Second, we include the lagged changes in a number of performance metrics as another set of instrumental variables (IVs): (i) the lagged category sales at the retailer, (ii) the lagged share of the focal brand at the retailer, and (iii) the lagged share of PLs in the category at the retailer.16 Third, we use a number of IVs to capture the evolution in overall production costs. To that extent, we account for changes in (i) the overall consumer price index, (ii) GDP growth and (iii) consumer confidence. Fourth, we account for systematic differences between groups of categories not yet captured by the previous sets of IVs. To that end, we include (i) the purchase frequency and penetration of the category (which capture the role of the category for the retailer see, e.g., Dhar et al., 2001), (ii) the concentration level in the category, and (iii) fixed-effect dummies for five category types (food, beverages, household care, personal care and a rest category). Finally, we also include fixed-effects dummies for each retailer, and a dummy variable to indicate whether the brand was also available at (competing) discounters.

Our instruments proved to be strong, as evidenced by the R²- and F-statistics: across the 30 regressions (6 marketing-mix instruments across 5 brand types), an average (median) R² of .54 (.57) was obtained, while all F-values exceeded the commonly-used threshold of 10. Moreover, Sargan tests firmly confirmed their validity: for the 30 instrumental variables in the 3 models, the null hypothesis of instrument exogeneity could never be rejected at any of the conventional significance levels.

4.4. Parameter estimates

No multi-collinearity problems are encountered, as the highest recorded VIF amounted to 5.4. We estimate our models for the four national supermarkets that represent the traditional platform for full

14 Similar summary statistics across all categories are available from the first author upon request. We do want to point out, however, that apart from the cross-sectional variation already revealed in Table 3, there is also considerable time variation in the data: calculating the coefficient of variation (CV) for each retailer–category–variable combination (6782 in total), we find that 2106 (31.1%) have a CV higher than .75, and 1676 (24.7%) a value that even exceeds 1.

15 Computed as the maximum difference between two competing brands in the category at the retailer.

16 If the focal brand is a PL, then this effect is already captured in the previous instrument (its own lagged share).
distribution coverage (but include the PL-dominated discount chains to compute the share and CDI variables). The CDI model estimates are based on 53 (categories) × 4 (key retailers) × 108 (periods) = 22,896 observations, whereas the retailer share model is based on 4 (key retailers) × 108 (periods) = 432 observations. Estimates of the manufacturer share parameters, in turn, are based on 5 (market players) × 4 (key retailers) × 108 (periods) × 53 (categories) = 114,480 observations. We estimate the market–share-atraction model using the log-centering procedure (Cooper & Nakanishi, 1988). To avoid over-parameterization of the manufacturer–share model, which contains intercepts for specific brands, we keep the category- and retailer-level deviations from those intercepts common across the non-Unilever brands, but allow them to be distinct from the Unilever brands. In all three models, we consider 6 marketing-mix variables, across 5 players (UL, NB1, NB2 and Other NBs). Across the three models, we consider 6 marketing-mix variables, across 5 players (UL, NB1, NB2 and Other NBs).

The parameter estimates used in the simulations for the focal, assortment-related, and constant-price-based dependent variables) amounting to .99 for the CDI model, .85 for the retailer share model, and .89 for the manufacturer-share model. The parameter estimates for the focal, assortment-related, and constant-price-based dependent variables) amounting to .99 for the CDI model, .85 for the retailer share model, and .89 for the manufacturer-share model.

Table 3
Summary statistics on the focal categories at the focal retailer Albert Heijn.*

<table>
<thead>
<tr>
<th></th>
<th>Bread</th>
<th>Cheese</th>
<th>Ice cream</th>
<th>Hot drinks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Assortment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># brands NB1</td>
<td>1.24</td>
<td>.39</td>
<td>1.26</td>
<td>.41</td>
</tr>
<tr>
<td># brands PL</td>
<td>2.63</td>
<td>.64</td>
<td>2.33</td>
<td>.45</td>
</tr>
<tr>
<td># brands UL</td>
<td>1.20</td>
<td>1.21</td>
<td>1.67</td>
<td>.52</td>
</tr>
<tr>
<td># sub-brands per brand</td>
<td>1.01</td>
<td>1.30</td>
<td>3.29</td>
<td>1.14</td>
</tr>
<tr>
<td>NB1</td>
<td>2.21</td>
<td>.25</td>
<td>1.28</td>
<td>.30</td>
</tr>
<tr>
<td>#sub-brands per brand PL</td>
<td>.99</td>
<td>.08</td>
<td>1.18</td>
<td>.30</td>
</tr>
<tr>
<td># SKUs per brand NB1</td>
<td>4.91</td>
<td>1.75</td>
<td>13.18</td>
<td>2.79</td>
</tr>
<tr>
<td># SKUs per brand PL</td>
<td>67.84</td>
<td>4.06</td>
<td>20.71</td>
<td>6.27</td>
</tr>
<tr>
<td># SKUs per brand UL</td>
<td>3.93</td>
<td>1.34</td>
<td>3.57</td>
<td>.86</td>
</tr>
<tr>
<td>Pricing/promotions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular price NB1</td>
<td>.25</td>
<td>.07</td>
<td>.96</td>
<td>.13</td>
</tr>
<tr>
<td>Regular PL</td>
<td>.26</td>
<td>.03</td>
<td>1.00</td>
<td>.07</td>
</tr>
<tr>
<td>Regular UL</td>
<td>.23</td>
<td>.04</td>
<td>1.61</td>
<td>.11</td>
</tr>
<tr>
<td>NB1 in feature (%)</td>
<td>.03</td>
<td>.08</td>
<td>.04</td>
<td>.14</td>
</tr>
<tr>
<td>PL in feature (%)</td>
<td>.59</td>
<td>.35</td>
<td>.34</td>
<td>.39</td>
</tr>
<tr>
<td>UL in feature (%)</td>
<td>.16</td>
<td>.27</td>
<td>.02</td>
<td>.07</td>
</tr>
<tr>
<td>% promo per brand NB1</td>
<td>.03</td>
<td>.06</td>
<td>.06</td>
<td>.06</td>
</tr>
<tr>
<td>% promo per brand PL</td>
<td>.03</td>
<td>.05</td>
<td>.02</td>
<td>.04</td>
</tr>
<tr>
<td>% promo per brand UL</td>
<td>.02</td>
<td>.03</td>
<td>.04</td>
<td>.04</td>
</tr>
</tbody>
</table>

* Similar summary statistics on (i) NI2 and the other NBs, as well as (ii) the 49 non-focal categories are available from the first author upon request.

5. Assessing the gains and losses of exclusivity

Having estimated the impact of the various marketing instruments (and especially those that capture distribution intensity) on category–retailer CDIs, retailer shares and manufacturer shares, we assess the sales implications of alternative distribution arrangements. As indicated before, we compare three scenarios: (i) a “null” scenario without the exclusive product line, (ii) an exclusive-distribution scenario, and (iii) an intensive-distribution scenario. We focus on the performance during the first year following the introduction, as is common in CPG studies on new-product acceptance (Gielens & Steenkamp, 2007). The predicted effects of the assortment changes depend on the estimated parameters, which have some inherent uncertainty. We therefore simulate the various performance metrics based on 500 parameter draws from the relevant normal distributions, and use the resulting empirical distribution to assess the significance of the effects.

All models were estimated using the commonly-available SAS routine Proc Mixed.
The foregone sales due to exclusivity therefore add up to €5520 K (i.e. from €3,652,790 K to €3,658,310 K) gain that would have been realized under intensive distribution. For Becel, intensive distribution would even have been detrimental to the focal retailer, as it lowered his sales from €6,152,908 K under the null scenario to €6,142,933 K.

In combination, these findings confirm both conventional wisdom and O'Brien and Shaffer (1997) analytical insights. However, there is considerable variation in the (relative) sizes of these gains and losses. Overall, a positive effect for the dyad can be found in three out of five cases (Boursin, Magnum and Lipton). In the other two cases, the dyad as a whole loses sales (Blue Band and Becel). Moreover, gains and losses need not be aligned. The largest retailer gain, for Becel, does not coincide with the largest manufacturer loss (for Blue Band), and the large retailer gain for Lipton comes along with one of the smaller manufacturer losses.

To further explore this, we decompose the total effects into (i) a focal- and other-category component and, for the manufacturer, into (ii) a focal- and other-retailer component. For the retailer (Table 6B), two points are worth noting. First, cross-category spillovers cannot be ignored. For instance, the introduction of Blue Band bread increases the focal retailer’s bread sales by €26,894 K (first term in Eq. (1)), but largely at the expense of other categories (−€19,550 K). Ignoring these cross-effects would lead Albert Heijn to evaluate the exclusive introduction as way too favorable. Second, cross-category effects can be negative as well as positive. For Becel and Blue Band, the retailer incurs negative cross-effects. This is consistent with Chen et al. (1999), who find that for a number of categories – including bakery products – the store-wide profits from merchandising (which they label ‘marketing profits’) are far lower than the within-category profit increases (labeled ‘accounting profits’). In other instances (i.e., Boursin, Magnum and Lipton), the retailer enjoys higher sales in non-focal categories – such positive cross-effects being the dominant (meta-analytic) pattern (p < .01).

As for the manufacturer (Table 6A), a large part of his losses can be attributed to foregone sales in the focal category at retailers excluded from the exclusive arrangement (e.g., for Blue Band: −€23,014 K). This loss is only partly compensated by increased manufacturer sales in the focal category at the focal retailer (+€2324 K). Cross-category spillovers also affect the performance implications for the manufacturer. Again, these cross-effects may be positive (e.g., for Magnum) as well as negative (e.g., for Blue Band and Becel), but with no clear overall pattern (p > .10).

The decomposition in Table 6A further reveals that even though the exclusive introduction of the two bread brands leads to a substantial gain for UL at the focal retailer (+€15,363 K for Blue Band, +€4047 K for Becel), the losses at excluded retailers are disproportionally higher (Blue Band: −€56,262 K, Becel: −€26,859 K). A possible explanation is that UL does not carry any bread-brand at those excluded retailers. Not entering the category with a first UL-brand at those retailers (as is the case with exclusive, as opposed to intensive, distribution) leads to huge opportunity losses. Opportunity losses are also incurred in the other cases, but to a lesser extent. As the introduction does not comprise an entirely new brand (but a sub-brand under an already available name), UL can still capture a share of consumers’ category purchases at the non-focal retailers through its available category-offer. Hence, manufacturers should carefully consider the brand architecture of their new and existing offering, not only with the exclusive retailer, but also with excluded ones.

5.1. Sales gains and losses of exclusivity

Table 5 reports, for each of the five cases, the total sales for the manufacturer (Panel A) and the focal retailer (Panel B) under the different scenarios. All figures refer to the total Dutch market (to translate the sample-based levels to population levels, we used conversion rates provided by GfK), and manufacturer gains and losses are calculated across the four national retailers.

In four out of the five cases, intensive distribution generates higher sales for the manufacturer than the exclusive arrangement. For instance, introducing Blue Band Bread exclusively increases the total yearly sales of Unilever from €736,327 K to €789,123 K; making the product available at all four national supermarkets would have yielded €830,022 K. The foregone sales due to exclusivity therefore add up to €40,899 K: almost half (44%) of the predicted sales increase under intensive distribution. Likewise, important losses occur for Lipton, Boursin and especially Becel–Magnum being the only case where the manufacturer does not lose from exclusivity.

In contrast, an exclusive arrangement is typically appealing for the retailer granted exclusivity. First, the exclusive deal leads to (substantially) higher sales than the null scenario, for each of the introductions. For Blue Band Bread, yearly sales at the focal retailer go up from €3,652,790 K to €3,665,654 K. In each of the introduction cases, this sales lift is far higher than that of a non-exclusive setting, such that the focal retailer gains from exclusivity. For Blue Band Bread, the €12,864 K sales lift under exclusivity clearly exceeds (p < .01) the

---

Table 4
Estimation results: marketing mix instrument estimates (means of the population mixing distributions).

<table>
<thead>
<tr>
<th>Effect</th>
<th>CDI</th>
<th>Retailer share</th>
<th>Manufacturer share</th>
</tr>
</thead>
<tbody>
<tr>
<td># brands</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL</td>
<td>0.022**</td>
<td>0.084**</td>
<td>0.136***</td>
</tr>
<tr>
<td>NB1</td>
<td>0.009</td>
<td>0.087**</td>
<td>0.350***</td>
</tr>
<tr>
<td>NB2</td>
<td>0.001</td>
<td>−0.197***</td>
<td>0.408***</td>
</tr>
<tr>
<td>PL</td>
<td>0.009</td>
<td>0.029***</td>
<td>0.405***</td>
</tr>
<tr>
<td>Other NBs</td>
<td>0.064***</td>
<td>0.069***</td>
<td>0.392***</td>
</tr>
<tr>
<td># sub-brands/brand</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL</td>
<td>0.015**</td>
<td>0.149***</td>
<td>0.036***</td>
</tr>
<tr>
<td>NB1</td>
<td>0.015**</td>
<td>−0.021</td>
<td>0.366***</td>
</tr>
<tr>
<td>NB2</td>
<td>−0.025**</td>
<td>−0.143***</td>
<td>0.192***</td>
</tr>
<tr>
<td>PL</td>
<td>0.034***</td>
<td>0.020***</td>
<td>0.099***</td>
</tr>
<tr>
<td>Other NBs</td>
<td>−0.310***</td>
<td>−0.036</td>
<td>0.181***</td>
</tr>
<tr>
<td># SKUs/brand</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL</td>
<td>0.014***</td>
<td>0.085***</td>
<td>0.111***</td>
</tr>
<tr>
<td>NB1</td>
<td>0.022**</td>
<td>0.002</td>
<td>0.268***</td>
</tr>
<tr>
<td>NB2</td>
<td>0.018***</td>
<td>0.137***</td>
<td>0.195***</td>
</tr>
<tr>
<td>PL</td>
<td>0.011**</td>
<td>0.003</td>
<td>0.359***</td>
</tr>
<tr>
<td>Other NBs</td>
<td>−0.108***</td>
<td>−0.190***</td>
<td>0.227***</td>
</tr>
<tr>
<td>Regular price</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL</td>
<td>−0.002</td>
<td>0.000</td>
<td>0.005***</td>
</tr>
<tr>
<td>NB1</td>
<td>0.000</td>
<td>−0.002</td>
<td>0.028***</td>
</tr>
<tr>
<td>NB2</td>
<td>−0.004</td>
<td>0.018***</td>
<td>0.003</td>
</tr>
<tr>
<td>PL</td>
<td>−0.002</td>
<td>−0.007**</td>
<td>0.003</td>
</tr>
<tr>
<td>Other NBs</td>
<td>−0.004*</td>
<td>0.008</td>
<td>−0.039***</td>
</tr>
<tr>
<td>Discount depth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL</td>
<td>0.002</td>
<td>0.014</td>
<td>0.128***</td>
</tr>
<tr>
<td>NB1</td>
<td>0.040*</td>
<td>−0.025</td>
<td>0.344***</td>
</tr>
<tr>
<td>NB2</td>
<td>0.014</td>
<td>0.010</td>
<td>0.164***</td>
</tr>
<tr>
<td>PL</td>
<td>−0.003</td>
<td>0.024**</td>
<td>0.175***</td>
</tr>
<tr>
<td>Other NBs</td>
<td>−0.001</td>
<td>−0.069**</td>
<td>0.395***</td>
</tr>
<tr>
<td>Feature</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL</td>
<td>0.010***</td>
<td>0.001</td>
<td>0.073***</td>
</tr>
<tr>
<td>NB1</td>
<td>0.000</td>
<td>0.045***</td>
<td>0.363***</td>
</tr>
<tr>
<td>NB2</td>
<td>0.001</td>
<td>−0.033**</td>
<td>0.178***</td>
</tr>
<tr>
<td>PL</td>
<td>−0.007**</td>
<td>−0.002</td>
<td>0.243***</td>
</tr>
<tr>
<td>Other NBs</td>
<td>0.013***</td>
<td>−0.008</td>
<td>0.125***</td>
</tr>
</tbody>
</table>

**p < .01; ***p < .05; **p < .10, based on two-sided t-tests.

---

21 As indicated by Chen et al. (1999), such negative cross-effects can arise when the consumers who are attracted to the store by the category spend little on anything else, or spend a lot on items that bring in little money for the retailer. We find this to be true for bread, where 35% of sales occur on fill-in trips (with below-median trip receipts), though these trips represent only 29% of total spending. By attracting more consumers on fill-in trips, total store sales may actually decline because of crowding (Lam, Vandenbosch, & Pearce, 1998), especially if large-basket customers are the ones that get crowded out (we thank an anonymous reviewer for this observation).
that reducing the offering of the leading competitor in the category (NB1, NB2, Other NBs and PL), we run structural-break tests on the feature-promotion time series indicate that it is not uncommon for the exclusive retailer to waive the trade-incentives for the step (for Magnum, list price and price-cut series at the focal retailer for the competing brand) and Verhoef and Sloot (2010) (who record negative store sales effects even if low-selling items are removed from the assortment) and the prospect of a significant increase in feature support for UL at the focal retailer (Magnum and Lipton: +10%-points; Becel: +30%-points).

5.2.2. Feature support

Instead of driving out competitive products, thereby weakening the appeal of the focal retailer, the manufacturer could attempt to negotiate better terms for bringing his own items to the store. Specifically, he can aim for enhanced support in the store flyer, ideally at the expense of the retailer. Indeed, informal exchanges with industry experts confirm that it is not uncommon for the exclusive retailer to waive the trade-promotion fees regularly charged for such feature support. Similar structural-break tests on the feature-promotion time series indicate that three of four exclusive introductions indeed coincided with a significant increase in feature support for UL at the focal retailer (Magnum and Lipton: +10%–points; Becel: +30%–points).

In line with the previous analyses, we subsequently consider the predicted sales gains and losses of the exclusivity arrangement under this increased support condition, and compare it to an intensive does not make the exclusive deal a winning proposition for the manufacturer. The competitive delisting does not systematically reduce the manufacturer exclusivity losses, nor does it make the exclusive deal profitable for the manufacturer in any of the three instances. Moreover, the delisting does not benefit the retailer and, if anything, dampens the retailer’s overall exclusivity gains (Magnum: from €3598 K to €8000 K, p < .01). Dropping these items may reduce the category’s total attractiveness at the focal chain (Mason, 1990) and, because of the lower category-draw effect (Campo, Gijbelsbrecht, Goossens, & Verhetsel, 2000), adversely affect the retailer’s total sales share — a finding in line with Borle et al. (2005) (who record negative store sales effects even if low-selling items are removed from the assortment) and Verhoef and Sloot (2010) (who find delisting to be particularly harmful in case of high-equity brands). Our calculations show that, indeed, the delisting leads to significantly lower retailer sales in the focal category (for Blue Band: in the bread category, p < .01) or overall (for Magnum, p < .01). The manufacturer, who fully depends on the focal retailer to sell his exclusive line, also suffers from this reduced category-draw effect, which dampens the advantages of lower category competition. Hence, even though the manufacturer may find the prospect of a reduced assortment for one of his leading competitors intuitively appealing, it may not produce the hoped-for effect (while potentially raising various antitrust issues).

5.2. Compensatory benefits for the manufacturer

Our results thus far suggest that, ceteris paribus, retailers stand to gain from exclusivity, while exclusive brand- or sub-brand introductions substantially reduce the manufacturer’s sales. The question then is why manufacturers would accept such arrangements. One explanation is that retailers are not willing to place the new product on the shelves, unless granted exclusivity. However, even in those circumstances, manufacturers may not be willing to carry the full load of the exclusivity losses, and try to negotiate some ‘quid pro quo’ to balance the risks/losses. It is well documented that retailers may negotiate slotting allowances, or impose fixed amounts of required trade support on the manufacturer, if they feel they carry too much of a new product’s failure risk (see, e.g., Sudhir & Rao, 2006). Likewise, manufacturers may request compensatory benefits from the focal retailer “for sacrificing part of their potential market when committing to an exclusive deal” (Cai et al., 2012, p. 172). We explore the consequences of three possible forms of compensation: (i) competitive delisting, (ii) increased feature support, and (iii) margin shifts.

5.2.1. Competitive de-listing

In return for exclusivity, the focal manufacturer may try to get some of his competitors delisted. To check for the occurrence of any incumbent assortment changes in our five cases, we run structural-break tests on the different assortment-related time series. Specifically, for each competitor in the category (NB1, NB2, Other NBs and PL), we regress his assortment variables (number of brands, number of sub-brands, and SKUs per brand) at the focal retailer against (i) the lagged value of this assortment variable, (ii) a pulse dummy (one at the time of the exclusive introduction, zero otherwise), (iii) a step dummy (zero before, and one following the exclusive introduction) and (iv) a deterministic trend.22 Significant negative coefficients for the step dummy then point to systematic assortment reductions.

While we do not observe any significant assortment reductions for the private label or secondary national brands, we do find evidence of assortment de-listing (i.e., a drop in the number of brands, sub-brands or SKUs) of the leading competitor for three introduction cases: Becel, Magnum and Boursin (see also last row of Table 1). Building on this finding, we re-analyze the impact of exclusivity for these three cases, by now comparing the predictions of (i) an intensive distribution scenario without incumbent assortment changes, to (ii) an exclusive arrangement in which the focal retailer implements the observed de-listings for NB1. Table 7A displays the results. Interestingly, it shows that reducing the offering of the leading competitor in the category

22 This presumes stationarity of the series at hand, which was established through the Levin, Lin, and Chu (2002) panel unit root test (detailed results are available from the first author upon request).
Table 6
Sales gains and losses of exclusive distribution: decomposition.¹ ² ³

<table>
<thead>
<tr>
<th>Blue Band</th>
<th>Boursin</th>
<th>Magnum</th>
<th>Lipton</th>
<th>Becel</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gain (loss) of exclusivity</strong></td>
<td><strong>Gain (loss) of exclusivity</strong></td>
<td><strong>Gain (loss) of exclusivity</strong></td>
<td><strong>Gain (loss) of exclusivity</strong></td>
<td><strong>Gain (loss) of exclusivity</strong></td>
<td><strong>Gain (loss) of exclusivity</strong></td>
</tr>
<tr>
<td>Total focal category</td>
<td>-20,690***</td>
<td>-162***</td>
<td>-72**</td>
<td>-222***</td>
<td>-14,130***</td>
</tr>
<tr>
<td>Focal category at focal retailer</td>
<td>2324**</td>
<td>11</td>
<td>33**</td>
<td>17</td>
<td>2002***</td>
</tr>
<tr>
<td>Focal category at other retailers</td>
<td>-23,014***</td>
<td>-173***</td>
<td>-105***</td>
<td>-239***</td>
<td>-16,132***</td>
</tr>
<tr>
<td>Other categories</td>
<td>-20,209***</td>
<td>-42</td>
<td>294**</td>
<td>-102***</td>
<td>-8882***</td>
</tr>
<tr>
<td>Total focal retailer</td>
<td>15,363***</td>
<td>168***</td>
<td>348***</td>
<td>536***</td>
<td>4047***</td>
</tr>
<tr>
<td>Other retailers</td>
<td>-56,262***</td>
<td>-372***</td>
<td>-126***</td>
<td>-860***</td>
<td>-26,859***</td>
</tr>
<tr>
<td>Gain (loss) of exclusivity</td>
<td>-40,899***</td>
<td>-204***</td>
<td>222**</td>
<td>-324***</td>
<td>-22,812***</td>
</tr>
</tbody>
</table>

Panel B: focal retailer

| Focal category at focal retailer | 26,894** | 585 | 90*** | 561 | 15,397*** | 0*** |
| Other categories at focal retailer | -19,550** | 2568*** | 3508*** | 10,060*** | -3433 | 0*** |
| Gain (loss) of exclusivity | 7344*** | 3153*** | 3598*** | 10,621*** | 11,964*** | 0*** |

¹: p < .01, **: p < .05 ;*: p < .10, based on a sign test across 500 draws from the parameter distributions.
²: All figures are in K Euros, refer to yearly values, and are extrapolated from the panel to the total Dutch market, based on conversion rates provided by GfK.
³: Significance levels are only reported for gains and losses.
⁴: Significance levels for overall direction of effect are based on Stouffer’s meta-analytic method of adding Z’s.

In the preceding sections, we focused on sales revenue. Ultimately, increase in feature support indicated between brackets.

Scenario without increased feature support. Table 7B summarizes the results. Tying exclusivity to extra appearance in the store flyer makes the deal (more) appealing for the manufacturer, and takes away the sales losses (e.g. Becel: from -€22,812 K, p < .01 to + €294 K, p > .10). The focal retailer’s sales gains do not increase significantly and, if the retailer has accepted to incur the ‘out-of-pocket’ costs of adding extra flyer space (and/or has waived the trade fees of this extra space typically charged to the manufacturer), this will lower his profit from the deal. Still, if this entices the manufacturer to accept the exclusivity deal, this may well be a worthwhile exercise.

5.2.3. Gross profit impact and margin negotiations

In the preceding sections, we focused on sales revenue. Ultimately, however, both the manufacturer and the retailer want to know how exclusive arrangements affect their bottom line (however, both the manufacturer and the retailer want to know how exclusive arrangements affect their bottom line). The manufacturer may be willing to accept sales losses if exclusivity allows him to re-negotiate the terms of the deal, which would make each unit sold more profitable. Calculating the gross profit gains and losses from exclusivity is straightforward if detailed unit margins are available for all manufacturers and retailers, across all brands (SKUs) and categories, under the intensive as well as the exclusive scenario. Obviously, such data would be impossible to obtain. What we do have is data on (i) the retailer margins (common across retailers) in all 53 categories (ranging between 6 and 51%), and (ii) margins of the focal manufacturer (by retailer) for the 35 categories in which he operates (ranging between 10 and 51%). This information is not perfect, for two reasons. First, it reflects average margins across SKUs in the category at each retailer. Second, it reflects the margins that apply in the actual distribution setting for these category SKUs, which might have been different under an alternative distribution scenario.

Still, keeping these limitations in mind, the information allows us to derive ‘ballpark’ gross-profit insights. For the five exclusive introductions, we calculate the gross profit that would result from either intensive or exclusive distribution. We then derive the corresponding gross-profit gain from exclusivity (i.e. the gross profit difference between the intensive and exclusive scenario) for the manufacturer, the focal retailer, and the dyad. The outcomes are reported in Table 8, and show a pattern similar to Table 5.

Table 7
Sales gains and losses of exclusive distribution with competitive de-listing or extra feature support.¹ ² ³

Panel A: impact of competitive de-listing

<table>
<thead>
<tr>
<th>Blue Band</th>
<th>Boursin</th>
<th>Magnum</th>
<th>Lipton</th>
<th>Becel</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gain (loss) of focal manufacturer</strong></td>
<td><strong>Gain (loss) of focal manufacturer</strong></td>
<td><strong>Gain (loss) of focal manufacturer</strong></td>
<td><strong>Gain (loss) of focal manufacturer</strong></td>
<td><strong>Gain (loss) of focal manufacturer</strong></td>
<td><strong>Gain (loss) of focal manufacturer</strong></td>
</tr>
<tr>
<td>No competitive de-listing (1)</td>
<td>-40,899***</td>
<td>-204*</td>
<td>222**</td>
<td>111*</td>
<td>-423</td>
</tr>
<tr>
<td>With competitive de-listing (2)</td>
<td>-41,910***</td>
<td>-93</td>
<td>201</td>
<td>111*</td>
<td>-423</td>
</tr>
<tr>
<td>Delisting effect on gain = (2)−(1)</td>
<td>-1011***</td>
<td>111*</td>
<td>-423</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel B: impact of extra feature support

<table>
<thead>
<tr>
<th>Magnum (1)³</th>
<th>Lipton (1)³</th>
<th>Becel (3)³</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gain (loss) of focal manufacturer</strong></td>
<td><strong>Gain (loss) of focal manufacturer</strong></td>
<td><strong>Gain (loss) of focal manufacturer</strong></td>
<td><strong>Gain (loss) of focal manufacturer</strong></td>
</tr>
<tr>
<td>No extra feature support (1)</td>
<td>222**</td>
<td>-324***</td>
<td>-22,812***</td>
</tr>
<tr>
<td>With extra feature support (2)</td>
<td>544***</td>
<td>-43</td>
<td>294</td>
</tr>
<tr>
<td>Feature effect on gain = (2)−(1)</td>
<td>322***</td>
<td>281***</td>
<td>23,106***</td>
</tr>
<tr>
<td><strong>Gain (loss) of focal retailer</strong></td>
<td><strong>Gain (loss) of focal retailer</strong></td>
<td><strong>Gain (loss) of focal retailer</strong></td>
<td><strong>Gain (loss) of focal retailer</strong></td>
</tr>
<tr>
<td>No extra feature support (1)</td>
<td>3598***</td>
<td>10,621***</td>
<td>11,964***</td>
</tr>
<tr>
<td>With extra feature support (2)</td>
<td>3665***</td>
<td>10,684***</td>
<td>12,058***</td>
</tr>
<tr>
<td>Feature effect on gain = (2)−(1)</td>
<td>67</td>
<td>63</td>
<td>94</td>
</tr>
</tbody>
</table>
In four out of five cases, the manufacturer incurs gross–profit losses, whereas the focal retailer always gains from the deal. However, these figures reflect the spread of gross–profit gains if the average manufacturer and retailer margins were to still apply under exclusive distribution. In reality, the exclusive arrangement is often bound to go along with margin-renegotiations (Trivedi, 1998). Discussions with industry participants confirm that this occurred for Blue Band Bread. The margins agreed upon by UL and AH under the exclusive deal indeed shifted part of the retailer’s gross–profit gains back to the manufacturer. An interesting take-away from Table 8 is whether such a margin shift could make the exclusive deal profitable for each of the parties involved. Focusing on the bottom row of the table, we observe that exclusivity leads to a significant gross–profit gain for the dyad in four out of five instances (Boursin, Magnum, Lipton and Becel). This suggests that, upon negotiating the exclusive contract, the manufacturer and focal retailer can arrive at a win–win situation: discussing margin compensations such that the party that would suffer from exclusivity at average margins manages to also benefit from the deal. This fits in with earlier analytical results that the prevalence of exclusivity deals crucially depends on the margin implications for the dyad as a whole (e.g., Trivedi, 1998), and that revenue sharing may turn the exclusive deal into an equilibrium (profitable) strategy for both members of the dyad (Cai et al., 2012).

For the four cases with a dyad gain, Table 9 shows the ‘break even’ margin shifts in the focal category, i.e. the transfer of margin from the retailer to the manufacturer, that would make the manufacturer (top row) or the retailer (bottom row) indifferent between the intensive and exclusive arrangement. The table shows that even if the manufacturer suffered from exclusivity under the base (average) margins, small shifts in margin within the focal category (e.g. 1.43 percentage points for Lipton, 3.18 percentage points for Becel) can already make the exclusive deal acceptable for the manufacturer, and make up for the initial exclusivity margin loss (including the cross-category loss). For the retailer, larger margin shifts are in order before the exclusivity arrangement would no longer be profitable (for example, –13.74 for Becel). In combination, these numbers give insights into the leeway that the channel partners have to arrive at a mutually acceptable win–win situation. For example, a shift of 5 percentage points would make an exclusive Becel deal profitable for the manufacturer (as this exceeds the minimal required shift of 3.18 to make him indifferent), while keeping the deal profitable for the retailer (as the reduction in his margin is smaller than the magnitude that would make him indifferent).

6. Conclusions

Even though they seem to go against what is commonly understood as ‘best practice’, exclusivity arrangements in grocery retailing are gaining way. This raises a number of compelling questions. Can exclusive deals for CPG lines be beneficial for both the manufacturer and the retailer, or only for one party? What drives these performance outcomes, and how can we guide retailers and manufacturers in designing such deals? To address these questions, we quantified the overall performance implications, and decomposed the gains and losses into a focal- and a cross-category (retailer) component. Next, we explored whether these negotiations in terms of incumbent-assortment, feature support, and margins make the exclusive deals more appealing to the manufacturer, the retailer, and/or the dyad. Several insights emerged.

First, the dominant picture is that exclusive arrangements enhance sales for the retailer granted exclusivity. Our results underscore that the retailer’s losses and gains from exclusivity should be calculated not only within the focal category, but also across other categories. Cross-category shifts make up an important part of the overall effect. Ignoring these shifts may substantially bias the anticipated outcomes of the deal. These cross-category effects may be positive (enhance the gains of carrying the exclusive line) or negative (lower the gains from the deal). For the retailer, it is important that requests for exclusivity be made in categories that entail positive spillovers.

The manufacturer, in contrast, typically realizes lower sales under an exclusive than under an intensive scenario, and thus incurs a loss of exclusivity. This corroborates that when it comes to sales, CPG manufacturers generally prefer intensive distribution. While category spillovers occur for the manufacturer as well, his exclusivity losses for the most part stem from foregone sales within the focal category at excluded retailers. Such losses are particularly high if the exclusive arrangement involves a category where the manufacturer was not yet present at those retailers. In such instances, the manufacturer cannot cater to the category needs of those excluded retailers’ customer base through his current offering. Still, even in categories where the manufacturer already captures a high share of the (focal and excluded) retailers’ assortment, exclusivity for new (sub-) brands tends to hurt sales.

So, why would manufacturers be willing to accept exclusive arrangements, and what guidance can we offer CPG manufacturers and retailers in setting up (or deciding whether or not to engage in) such deals? Our results suggest that, to create win–win outcomes, manufacturers should not try to have leading incumbent products de-listed: this is not likely to render exclusivity beneficial to them, nor is it in the interest of the exclusive retailer. In contrast, extra feature support for the exclusive brand may make an otherwise harmful arrangement beneficial for the manufacturer. Whatever the sales gains or losses, manufacturers and retailers want to know how exclusivity affects their bottom line. To further explore this issue, we calculated the combined gross–profit implications for the retailer–manufacturer dyad. We find that four out

---

**Table 8**

Gross profit implications of exclusive distribution, based on average margins.a,b

<table>
<thead>
<tr>
<th></th>
<th>Blue Band</th>
<th>Boursin</th>
<th>Magnum</th>
<th>Lipton</th>
<th>Becel</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focal retailer</td>
<td>–7355**</td>
<td>–57*</td>
<td>69***</td>
<td>–122**</td>
<td>–901**</td>
<td>&lt;0***</td>
</tr>
<tr>
<td>Dyad</td>
<td>6373**</td>
<td>1163**</td>
<td>1130**</td>
<td>3461**</td>
<td>3893**</td>
<td>&gt;0***</td>
</tr>
</tbody>
</table>

**Table 9**

Break-even margin shifts.a

<table>
<thead>
<tr>
<th></th>
<th>Blue Band</th>
<th>Boursin</th>
<th>Magnum</th>
<th>Lipton</th>
<th>Becel</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BE retailer margin</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Margin shift</td>
<td>n.a.</td>
<td>–2.30</td>
<td>–0.6</td>
<td>1.43</td>
<td>3.18</td>
</tr>
<tr>
<td><strong>BE manufacturer margin</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Margin shift</td>
<td>n.a.</td>
<td>–46.61</td>
<td>–9.95</td>
<td>–40.38</td>
<td>–13.74</td>
</tr>
</tbody>
</table>

---

a All figures are expressed in K Euros, refer to yearly values, and are extrapolated from the panel to the total Dutch market, based on conversion rates provided by CBK. UL gross profit is calculated across the four key chains.

b For Magnum, Ola, Lipton and Boursin, average margins for the retailer and the focal manufacturer are directly obtained from the data provider. For Becel and Blue Band, average manufacturer margins are obtained by deducting average retail margins in the category from total dyad margins in the category (figures, again, obtained from our data provider).

c Margin shifts in the focal category, i.e. the transfer of margin from the retailer to the manufacturer, that would make the manufacturer (top row) or the retailer (bottom row) indifferent between the intensive and exclusive arrangement.

---

25 These retailer gains are of a similar order of magnitude as reported in ter Braak, Dekimpe, and Geyserkens (2013) in the context of contract negotiations with PL suppliers, which they refer to as a “substantial absolute improvement in retailer profitability” (p. 97, italics added).

26 No such (behind-the-scene) insights were available for the other cases.

27 For Becel, exclusivity leads to a profit gain for the dyad (Table 6), in spite of a sales loss (Table 5). This is attributable to the fact that (for this category) the margin of the party which experiences a sales gain (the retailer) is much higher than the margin made by the party experiencing a sales loss (the manufacturer).
of five actual exclusive introduction cases yielded a gross profit gain for the dyad as a whole, which is in line with the analytical results of Cai et al. (2012) in the telecom industry.

In such situations, margin re-negotiations may result in a win–win outcome for each of the parties involved. This would necessitate a margin shift from the retailer toward the manufacturer to compensate the manufacturer for his sales losses from the exclusive deal. Whether the manufacturer will succeed in obtaining such margin shift will depend on the parties’ relative power: larger manufacturers may incur higher absolute sales losses, but may be better placed to obtain margin compensations. Another caveat is that to maintain good relationships with excluded retailers, the manufacturer may have to offer them something in return as well. Though our expert-contacts did not elaborate on the nature of these compensations, our data seem to point to increased feature activity by the focal manufacturer at excluded retailers following the exclusive deal (break tests reveal systematic increases in Unilever’s feature activity in the focal category at competing retailers, in 5 out of 9 cases).20 Even if this compensation enhances sales of the focal manufacturer, the increased trade investment is likely to reduce his profitability. The focal retailer, in turn, will see his sales and profit go down. Hence, to the extent that exclusive arrangements call for compensations of non-focal retailers by the focal manufacturer, this will reduce the dyad’s ‘room for maneuvering’.

6.1. Limitations/future research

Several areas for future research remain. Even though we built on a long tradition of using restricted policy simulations, and even though we made sure to stay within the range observed in our data, one could argue that our results are susceptible to the Lucas critique (see, e.g., van Heerde et al., 2005). We could expand our analyses with super-exogeneity tests as used in van Heerde et al. (2007, 2010). However, the sheer number of possible effects (given that we work with 6 instruments of 5 players across 4 retailers and 3 different models) would make this impractical, and subject to the well-known multiple-testing problem. More research is needed on how to cope with the Lucas critique in such a high-dimensional setting. Relatedly, we focused in our estimation on the four largest national supermarkets. Also using smaller (or regional) chains and/or hard discounters21 would have resulted in more cross-retailer heterogeneity, and a more inclusive picture of the cross-retailer gains/losses, but would have considerably increased the dimensionality of the problem.

Also substantively, several areas for future research remain. First, while we found the direction of many of the effects to be consistent across the five introduction cases, the size of the exclusivity gains and losses differed. As evidence on more exclusive deals becomes available, future studies could investigate how the exclusive line’s brand architecture, price position and qualitative fit with the rest of the category affect the performance implications. While our current analysis quantifies the performance implications of existing exclusivity arrangements, the availability of such a contingent framework would allow to also evaluate a priori proposed deals.

Second, though we analyzed how exclusive introductions affected the presence of items in the retail assortment, we could not – for lack of data – study how they altered the amount or quality of shelf space for these items — something that we leave for future study. Similarly, while we considered three aspects on which manufacturers could re-negotiate the terms-of-the-deal (i.e., competitive de-listing, feature support and margin shifts), other options remain. For example, it would be interesting to verify whether exclusive deals go along with changes in wholesale prices (Luo, Kannan, & Ratchford, 2007), slotting allowances (Sudhir & Rao, 2006) or trade-support requirements (Bord Bia, 2010), and/or with shifts in pass-through rate between the manufacturer and the retailer — changes that could also affect the bottom-line implications of exclusive deals.

Third, we focused on unique product lines, i.e. where the exclusivity agreement applied to an entire range of SKUs. Other retail-focused customization efforts deal with exclusive SKUs, ranging from unique formulations and sizes to packaging options. As such, Ricola, the Swiss manufacturer of cough drops, offers a lemon verbena flavor uniquely to Whole Foods shoppers in the U.S. market. It may be useful to investigate when exclusive brands/product lines may be preferred over exclusive SKUs. Also, exclusivity need not be permanent. As such, U.K. retailer Asda secured exclusivity for P&G’s British launch of its “Herbal Essence Fruit Fusions” line for three months, after which the product was rolled out to other retailers. Little is known about these temporary exclusivity deals, and more research is needed.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ijresmar.2014.03.003.

Acknowledgments

The authors gratefully acknowledge Sjoerd Schaafsma (Unilever) and Alfred Dijis (A@Mark) for providing the data and for sharing their expertise. They also thank Jan-Benedict Steenkamp, Harald van Heerde, Rik Pieters, and the participants of the Tilburg Xmas Camp, the MSI Conference on Marketing Metrics for the Connected Organization, and the Marketing Seminar Series at UNC for their helpful comments and suggestions.

References


How to protect your premium product from low-price competitors: Price, quality, or portfolio adjustment?

Peter-J. Jost 1

Institute for Organization Theory, WHU — Otto-Beisheim-School of Management, Burgplatz 2, D-56179 Vallendar, Germany

A R T I C L E   I N F O
Article info
Article history:
First received in May 18, 2010 and was under review for 1-1½ months
Available online 12 March 2014
Area Editor: David Soberman

A B S T R A C T
In a game-theoretic framework, I analyze how a brand manufacturer can thwart new entrants into its market. Three strategic options are considered, a price adjustment of the premium product, a quality adjustment of the premium product and a portfolio adjustment of adding a fighter brand. In a basic setup, I show that the incumbent’s best response to entry is to choose a portfolio adjustment. If, however, the incumbent is uncertain about whether the rival firm will enter the market, a price adjustment of the premium product might be the better alternative if launching the fighter brand is associated with costs. Moreover, if technological progress improves the efficiency of product development, a combined quality and portfolio adjustment might be the best alternative for the incumbent.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Brands are one of the few strategic assets available to a company that can provide a long-term sustainable competitive advantage. Wrigley’s, Coca-Cola or Gillette are examples for such brands that have remained market leaders over more than 80 years. These brand manufacturers enjoy higher-than-average margins because consumers are willing to pay high prices for their products. However, keeping the price level high gives rise to increasing competition from competitors entering the market with aggressive pricing. Generics, private labels or “clones” imitating the brand leaders are such low-price competitors. They enter markets where patents protecting the leader’s technologies have run out, where the market volume, market growth or the brand’s position is so labor expensive that foreign manufacturers have strong cost-advantages.

What should a brand manufacturer do to counter such attacks? The marketing literature on brand management suggests several reactions, see, for example, Aaker (2004, p. 230ff) or Keller (2008, p.224ff). Besides not to react at all there are three main types of basic responses to such low-price competitors:

• Price adjustment: One of the traditional reaction of a brand manufacturer is to lower the price of its premium product significantly. In the consumer good industry major brands usually cut prices to compete with private labels. For example, StarKist reduced the prices of its tuna to only five cents above the price of private labels. As a consequence, the share of private labels in this category sliced in half from 20% to 10%, see Keller (2008, p.224).

• Quality adjustment: To stay ahead of low-price competitors brand manufacturers may also emphasize innovation to improve existing brand products. Using this response requires a reposition of the premium product. H.J. Heinz, for example, has retained more than 50% market share in the ketchup category for years by aggressive packing and product developments, see Keller (2008, p.224).

• Portfolio adjustment: A third option which has not been used widely so far, is to enlarge the product portfolio. In this case the brand manufacturer adds a second, lower positioned product, a so called “fighter brand”, to the existing higher positioned brand product. For example, Philip Morris has used Chesterfield as a fighter brand to flank its brand Marlboro, see Quelch and Harding (1996, p.106).

Of course, these general responses to fight against new entrants could also be combined and modified. An example is Procter & Gamble with its leading diaper brand Pampers, see Berry and Schiller (1994). As the market share of private labels grew P&G repositioned its number three brand, Luvs, as a fighter brand. Despite its efforts to ensure that Luvs offers considerably less value than Pampers, Luvs stole sales from the premium brand. Only after a quality adjustment of the premium brand in form of a new, thinner diaper did Pampers recover.

This shows like many other case studies in the literature that neither a price nor a quality nor a portfolio adjustment is always successful. In fact, for each of these responses there exist several counterexamples demonstrating that this particular instrument failed for the company considered: Philip Morris’ decision, for example, to defend the market position of its brand Marlboro in 1992 by offering a 20% price cut per pack resulted in a 15% decrease in total revenues, see Hoch (1996). Or
United Airlines, after launching its fighter brand Ted in 2004 to fight against discount airlines, ceased operations in 2009 after years of failure, see Ritson (2009). It is for this reason, that some researchers advise to avoid fighter brands at all, see Jain and Tucker (1997), while other researchers recommend exactly the opposite, e.g. Ritson (2009), or to use other responses instead, for example Steiner (2004, p.118), for whom innovation is “one of the strongest competitive weapons”.

In this paper, I propose a model that highlights the costs and benefits of these options and evaluates the best response of a brand manufacturer in a game-theoretic framework. In particular, I consider a vertically differentiated market in which an incumbent firm can offer products of different qualities over two periods. In the first period the incumbent acts as monopolist whereas the market becomes a duopoly in the second period. This two-period framework allows me to analyze not only the short-run, one period consequences of entry into this market, but also the long-run effects potential entry has on the monopolistic behavior of the incumbent firm. To concentrate on the incumbent’s optimal response to entry I assume that consumers are uniformly distributed according to their willingness to pay for quality. This assumption implies that the market has one heterogeneous market segment only. In particular, the monopolistic firm never finds it optimal to produce a second brand without the threat of competition as a consequence of a decrease in its marginal revenues in the market. Within this framework, I then analyze the three above mentioned different basic responses of the incumbent firm to entry.

In the first scenario, the incumbent adjusts the price of its premium product in response to entry. This price reduction implies that the price-quality ratio of its premium product shrinks so that the market demand for its product is higher in the second period than in the monopoly period. On the other hand, however, a price cut implies that contributions are lost from those customers that previously had been willing to pay the higher price for the premium product.

In a second scenario, I consider the case in which the incumbent adjusts the quality of its premium product. Such a quality adjustment can go into two distinct directions. One possibility for the incumbent is to decrease its product quality. However, this implies that price competition becomes tougher and, in turn, leads to less profits. The other possibility is to upgrade its product quality. Although this direction of quality adjustment softens price competition and thus implies higher profits, improving product quality comes with costs for the incumbent. Since the incumbent lost its monopoly position, its marginal benefits from quality improvement are lower in the second than in the first period.

In a third scenario, I assume that the incumbent firm extends its product portfolio by offering a fighter brand. The challenge here is to position the second product such that it serves two purposes: On the one hand, it should fend off the entrant in order to directly compete with the premium product. On the other hand, the introduction of the second brand should not lead to cannibalization in the sense that current consumers of the premium product switch to buy the incumbent’s lower quality product although they would have never switched to the low-price entrant’s product. To reduce this negative effect of cannibalization the incumbent then tries to position the quality level of its second product as low as possible to soften price competition. This, however, irretrievably leads to a situation in which the competitor would launch its product in the middle market segment, thus implying more price competition in the market. Trading off these effects I show how the incumbent can optimally position its second product such that the upper half of the entire market is still covered by its premium product.

Finally, I compare the three possible responses of the incumbent and analyze under which circumstances which option is more beneficial. In the basic model it turns out that a portfolio adjustment is always the best alternative. If, however, launching a second product implies a large enough cost for the incumbent, a price adjustment becomes the more valuable the greater the incumbent’s uncertainty is about whether the rival firm will actually enter the market or not. Moreover a quality adjustment becomes beneficial if the incumbent’s investments’ efficiency improves from Period 1 to Period 2. In particular, the introduction of a fighter brand accompanied by a quality improvement of its premium brand is then the best option for the incumbent firm as a response to entry.

The present paper is organized as follows. The next section presents the basic model of monopoly in a vertically differentiated market. In Section 3 entry into this market is considered in a basic model and the three strategic options of the incumbent to thwart the entrant are analyzed and compared. Section 3.5 then extents this basic model in three directions and considers the effect of imitation costs by the entrant, launching costs by the incumbent and the possibility of technological progress on equilibrium behavior and the incumbent’s optimal response. Section 4 then concludes with some final remarks. Most proofs of the results are presented in Appendix A.

2. Monopoly in a vertically differentiated market

As a benchmark I briefly study in this section a model of vertical product differentiation under monopoly, see Gabszewicz and Thissen (1979), Shaked and Sutton (1982) and Choi and Shin (1992). First, the preferences of consumers are described and the supply side is introduced. I then analyze the monopoly case in the absence of entry which is later compared with the entry game.

2.1. The basic framework

Consider a market in which an incumbent firm I can offer products in different qualities over two periods. Assume that consumers can be described by a parameter θ which is uniformly distributed on the interval [0,1] with unit density, where θ ≥ 1. The parameter θi of consumer i can be interpreted as her willingness to pay for a product. Each consumer buys not more than one unit of the product per period of the qualities available in the market place. Her net surplus when buying a product with quality q at price p then is

\[ u_i(q, p) = q\theta_i - p. \]  

(1)

Consumers who do not purchase receive zero utility. I assume that (1) holds for all consumers in both periods.

On the supply side, the incumbent introduces its products in the beginning of Period 1 to the market. To offer a variety of product qualities \( q_1, ..., q_n \) with \( q_1 > ... > q_n > 0 \) for \( n \geq 1 \), positive development costs are required. I assume that the product with the highest quality determines total development costs. In particular, the incumbent has to incur development costs \( c(q) \) to produce the highest quality q = \( q_1 \) with

\[ c(q) = \frac{1}{2} \gamma q^2. \]  

(2)

where \( \gamma > 0 \) is a parameter that reflects, for example, the efficiency of the incumbent’s investments. When choosing its other product qualities \( q_2, ..., q_n \), the incumbent then has no further development costs such that it can costlessly develop products of lower quality. Moreover, I assume that the variable costs of production are independent of quality and equal to zero.

The incumbent offers its product qualities at monopoly prices \( p_1^0, ..., p_n^0 \) in both periods \( t = 1, 2 \). If \( x_i^t \) denotes the resulting demand of product \( (q_1, p_i^0) \), \( i = 1, ..., n \), the incumbent’s monopoly gross profits over both periods are

\[ \pi_d = \sum_{t=1}^{n} p_i^0 x_i^t + \delta \sum_{t=1}^{n} p_i^0 x_i^t - c(q_1). \]

where profits in Period 2 are discounted by the factor \( \delta \in [0, 1] \).
2.2. Product differentiation in the monopoly case

As a benchmark consider briefly this monopoly case. So suppose that there is no potential entrant and the incumbent acts as a monopolist over both periods. Of course, monopoly prices are identical in both periods as well as demand.

Consider the case in which the monopolist offers only one product \((q_{0},p_{M})\) and let \(\theta_{M} \in [0,\bar{\theta}]\) be the marginal consumer who is served by the monopolist. Then \(\theta_{M}\) is given by the condition \(q_{0}\theta_{M} - p_{M} = 0\), that is, \(\theta_{M} = p_{M}/q_{0}\). Hence, all consumers with a higher willingness to pay than \(\theta_{M}\) buy \((q_{0},p_{M})\) and demand is \(x_{M} = (\bar{\theta} - \theta_{M})\). The monopoly price then maximizes \(p_{M} (\bar{\theta} - \theta_{M})\) and is characterized by the condition that the marginal gain from serving an additional consumer is equal to the loss of all other consumers. Using the optimal monopoly price as a function of quality \(q_{0}\), the incubent then maximizes profits.

**Proposition 1.** In the absence of entry, the monopolist offers the following price-quality combination in both periods

\[
q_{M} = \frac{\varphi (1 + \delta)}{4\gamma_{1}}, \quad p_{M} = \frac{\varphi (1 + \delta)}{8\gamma_{1}}.
\]

Moreover, the incubent never has an interest to produce more than this product quality.

The optimal monopoly price-quality relation always splits the market equally, \(\theta_{M} = \bar{\theta}/2\): Only those consumers whose willingness to pay this price is higher than the one of the medium consumer will buy the product. As a consequence, the monopolist's revenues are increasing in product quality due to a higher monopoly price, and, in turn, since quality is costly, the optimal product design balances marginal revenues and marginal development costs.

Since a high quality product only captures the upper part of the market, it seems beneficial for the monopolist to introduce a second lower quality product. However, although the monopolist can enlarge the overall demand in the sense that some consumers who didn't buy a product before will now buy the lower quality product, the cannibalization effect outweighs this demand effect: In fact, some consumers now switch from the high quality product to the low quality product. Since the marginal gain from selling the high quality product is always greater than the marginal gain from selling the low quality product, the monopolist has an incentive to increase the low quality product price up to the point at which the price-quality relations of the two products are identical. But then nobody would demand the low quality product and, consequently, the monopolist would not introduce a second product at all.\(^2\)

### 3. Entry into a vertically differentiated market

In this section I extend the model of vertical product differentiation in Section 2 and consider entry in Period 2. According to the discussion in the introductory section, one can distinguish between three different responses of the incumbent firm to entry, see Fig. 1: First, the incumbent adjusts its price setting behavior; second, the incumbent adjusts also the quality of its premium product; and third, the incumbent firm extends its product portfolio by offering a fighter brand.

In the following Section 3.1, I first describe the basic framework of entry in Period 2 assuming that the incumbent firm does not adjust its initial product. This case of price adjustment as a response to entry is analyzed in Section 3.2. I then consider the incumbent's option to adjust the quality level of its initial product for Period 2, the case of a quality adjustment. In Section 3.4 I then assume that the incumbent introduces a second product, the case of portfolio adjustment. These three different responses of the incumbent firm to entry are then compared in Section 3.5.

#### 3.1. The basic framework

I extend the previous model of vertical differentiation as follows: In the beginning of Period 1, the incumbent introduces a product with quality \(q_{E} > 0\) in the market and remains monopolist during this period. In the beginning of Period 2, a second firm, called the entrant, enters the market. The entrant then competes for consumers by offering a product with quality \(q_{F} > 0\) when choosing its product quality \(q_{F}\) I assume that the entrant's cost function captures two stylized facts relevant in business practice. First, an entrant usually has the possibility to imitate the incumbent's products, and, therefore, to obtain a certain quality level \(q_{E} \leq q_{F}\) with lower development costs. I assume here that there is full imitation, that is, the entrant has free access to the incumbent’s technology and, hence, no development costs for a lower quality level. And second, the entrant as quality follower usually has a lower level of efficiency than the incumbent to develop a higher level of quality \(q_{E} > q_{F}\). Both assumptions guarantee that the incumbent is always the high quality firm and the entrant produces a lower quality as follower.

The incumbent observes the entrant’s quality choice \(q_{F}\) and both firms then set prices \(p_{E}\) and \(p_{F}\) conditional on \(q_{E}\) and \(q_{F}\). This choice then determines firms’ demands \(x_{E}^{F}\) and \(x_{F}^{E}\) in the second period. Since production is costless the entrant’s profits in Period 2 are

\[
\pi_{E} = p_{E}^{2}x_{E}^{F}.
\]

The total profits of the incumbent over both periods then are

\[
\pi_{I} = \pi_{I}^{1} + \delta\pi_{I}^{1} - c_{I}(q_{I}) = p_{I}^{1}x_{I}^{1} + \delta p_{I}^{1}x_{I}^{1} - c_{I}(q_{I}).
\]

In sum, I consider a two-period model with the following sequence of events:

- **Stage 1a** The incumbent sets a quality level \(q_{E}\) and charges a price \(p_{I}^{1}\) for the product in the first period.
- **Stage 1b** Consumers choose whether to purchase the product or not.
- **Stage 2a** A second firm chooses a product of quality \(q_{F}\) and enters the market.
- **Stage 2b** Having observed the product qualities offered in the market, the two firms compete by simultaneously choosing prices \(p_{E}^{1}\) resp. \(p_{F}^{1}\) in the product market.
- **Stage 2c** Finally, a consumer will buy from the firm that offers the best price-quality combination or she doesn’t buy at all.

Whereas this game structure assumes that the incumbent firm foresees potential entry in Period 2 when choosing its product quality in Period 1, an alternative scenario is possible in which the incumbent’s...
premium quality brand is fixed and the game starts in Period 2. Since the price–quality ratios as well as the products’ demands do not change compared to the general two-period model I focus in the following on the long-term perspective and include the first period into the analysis.

3.2. Option 1: price adjustment

To analyze firms’ optimal behavior, I use backward induction and start with the consumers’ consumption decision in Stage 2c. For a given quality level \( q_1 \) of the incumbent firm suppose that the second firm entered the market and offered a quality choice \( q_{11} < q_1 \). Then the price setting behavior of the two firms in Stage 2b as well as the entrant’s quality decision in Stage 2a differs in no way from the model of Choi and Shin (1992) and leads to the well known result that \( q_{11} = \frac{3}{4} q_1 \). When introducing product quality \( q_1 \) in Stage 1a, the incumbent then maximizes total pro

\[
\pi_{11} = \frac{1}{4} P_{11}^2 q_1 + \theta^2 \frac{7}{48} q_1 - \frac{1}{2} \gamma q_{11}^2.
\]

From the first-order condition one can then immediately derive the optimal quality level \( q_{11}^* \).

**Proposition 2.** Under price adjustment the incumbent responds to entry with lowering its product price under the first period monopoly price. Its product quality is below the monopoly quality but still higher than the entrant’s one.

**Proof.** See Appendix A.

Several remarks are worth noting: First, under entry the incumbent firm still offers the high quality product and leaves the lower part of the market to the entrant. However, since the incumbent foresees entry in Period 2, it chooses a lower product quality than in the monopoly case to defend its position:

\[ q_{1M} > q_{11} = \frac{P_{11}(12 + 7\theta)}{48\gamma_1} > q_{31} = \frac{P_{11}(12 + 7\theta)}{84\gamma_1}. \]

Product prices are ordered according to these qualities with the following properties: In the first period, the incumbent covers the upper part of the market as in the monopoly case. Price competition in the second period, however, implies that the incumbent firm uses price cuts to defend the rival’s entry in Period 2. In turn, the price–quality ratio shrinks so that more consumers with a lower willingness to pay now buy the high quality product:

\[
\frac{P_{1M}}{q_{1M}} = \frac{P_{11}}{q_{11}} = \frac{1}{2} \frac{\theta}{\gamma} > \frac{P_{11}}{q_{11}} = \frac{1}{4} \frac{\theta}{\gamma} > \frac{P_{11}}{q_{11}} = \frac{1}{8} \frac{\theta}{\gamma}.
\]

This also reflects market demand which is higher for product \( q_1 \) in Period 2 than in the monopoly case,

\[ x_{1M} = \frac{1}{2} \frac{\theta}{\gamma} x_{11} > \frac{7}{12} \frac{\theta}{\gamma} x_{11} = \frac{7}{24} \frac{\theta}{\gamma}. \]

Overall, the incumbent’s profit is lower than its monopoly profits but higher than the entrant’s ones.

3.3. Option 2: quality adjustment

Besides a price adjustment the incumbent could also consider an adjustment of its product quality. Two directions are possible: The incumbent could either decrease product quality or, alternatively, upgrade its product.

Consider the first alternative and suppose the incumbent lowers its product quality from \( q_{11} \) to \( q_{12} \) before the second firm enters the market. From the model of Choi and Shin (1992) it is well known that the entrant positions its product in a low quality range by setting \( q_{12} = \frac{3}{4} q_{11} \).

This, however, implies that the equilibrium prices are increasing in the product quality the incumbent offers,

\[
p_{12} = \frac{1}{4} q_{12}^2, \quad \text{and} \quad p_{22} = \frac{1}{14} q_{12}^2.
\]

Hence, price competition becomes tougher the more the incumbent adjusts its high product quality downwards and, in turn, leads to less profit. In sum, the incumbent has no incentive to adjust the quality of its premium product downwards.

However, this argumentation suggests, that it might be beneficial for the incumbent to adjust the quality of its premium brand upwards as a response to potential entry. Concerning price competition this implies higher prices and, in turn, higher profits for the incumbent. To discuss this in more detail suppose that the incumbent firm can upgrade its product \( q_1 \) before the second firm actually enters the market. Of course, improving product quality comes with positive costs for the incumbent.

Using the cost function (2) for the development of product quality I assume in the following that a quality improvement to \( q_{12} > q_{11} \) implies costs

\[
c(q_{12}) = \frac{1}{2} \gamma_1 \left( q_{12}^2 - q_{11}^2 \right).
\]

Hence, the incumbent’s development costs \( \gamma q_{12} / 2 \) from the first period are sunk and any later quality improvement in Period 2 is as expensive as it would have been in Period 1. This implies that the incumbent’s marginal costs of quality improvement are identical to the ones in Period 1, \( c'(q_{12}) = \frac{1}{2} \gamma_1 \gamma_1 \). On the other hand, however, the incumbent’s profits in Period 2 are lower than the ones in Period 1 due to competition,

\[
\pi_{22}(q_{12}) = \frac{1}{4} q_{12}^2 - \frac{1}{2} \gamma_2 q_{12} > \pi_{21}(q_{11}) = \frac{7}{38} q_{11}.
\]

This not only implies that marginal benefits from quality improvement are lower in the second than in the first period, but also implies that the incumbent has no incentives to improve the initial quality level of its product.

**Proposition 3.** In the basic model, it is never optimal for the incumbent to respond to entry with quality adjustment.

3.4. Option 3: portfolio adjustment

A third option for the incumbent is to adjust its product portfolio as a response to entry. Under this option, the incumbent firm has the possibility to introduce a second product to thwart the new entrant. Of course, the previous discussion suggests that this second product has to be of lower quality than the initial product \( q_1 \). Indeed, positioning a second product higher in quality than the existing product cannot be optimal for the incumbent due to additional marginal costs but reduced marginal benefits in Period 2 compared to Period 1. As a consequence, I term the initial product as the high quality brand \( q_{1M} \) and the second, new product as its lower quality product

\footnote{In the absence of entry costs, this assumption is not restrictive. Indeed, knowing that the second firm always enters the market, it is always beneficial for the incumbent to improve the quality of its product ex ante before entry occurs than ex post after entry occurred. This is because in the first case the incumbent has the possibility to influence the entrant’s choice of product specification whereas the incumbent would only react to entry in the second case.}
q_{1t}. The sequence of events in this modified interaction then is as follows:

Stage 1a The incumbent sets a quality level q_{1t} and charges a price p_{1t} for the product in the first period.
Stage 1b Consumers choose whether to purchase the product or not.
Stage 1c The incumbent has the possibility to credibly introduce a second product quality q_{2} with q_{2} < q_{1t}.
Stage 2a A second firm enters the market and offers a product of quality q_{3t}.
Stage 2b Having observed the product qualities offered in the market, the two firms compete by simultaneously choosing prices \( p_{1}^{*}, p_{2}^{*}, p_{3}^{*} \) in the product market.
Stage 2c Finally, a consumer will buy from the firm that offers the best price–quality combination for her or she doesn’t buy at all.

This game has the following solution:

**Proposition 4.** When responding to entry with portfolio adjustment the incumbent positions its fighter brand such that its premium product still covers the upper part of the market while the entrant launches a lower quality product.

**Proof.** See Appendix A.

In equilibrium, the incumbent uses its second product as a firewall in order to weaken the market position of the entrant, that is

\[
q_{2t} = \frac{\theta^2 (12 + (12 - 5\hat{p}t))}{48Y_{t}} > q_{1t} - \beta q_{1t} > q_{3t} = \frac{4}{7}q_{1t},
\]

with \( \beta = 0.54806 \). By positioning its second product in its quality below the one of the top product the incumbent fends off the rival firm in order to directly compete with its premium product. Of course, this introduction of a second product has drawbacks for the incumbent’s overall profits because it still competes with its premium product. Such cannibalization implies that current consumers of the premium product switch to buy the new second product although they would have never switched to the rival’s low-price product. To reduce this negative effect of cannibalization the incumbent tries to position the quality level of its second product as low as possible to soften price competition with the rival’s product, and in turn, to soften price competition between its own products. However, if the quality level of the second product was too low, it would lose its purpose as a firewall and the competitor would launch its product in the middle market segment. Although this would eliminate any cannibalization, such a positioning would lead to more price competition in the market. As a result, this not only reduces the demand for the incumbent’s second product but also for its premium product. Using the fighter brand as a firewall, however, allows the incumbent to still cover the upper half of the entire market by optimally choosing its product prices, as in the first period,

\[
x_{1t}^{*} = x_{2t}^{*} = \frac{\theta}{2}.
\]

### 3.5. The optimal response

If one compares the incumbent’s price adjustment with its portfolio adjustment, the latter one has two advantages: First, the incumbent can set a higher quality product in Period 1, that is, \( q_{2t} > q_{3t} \), see Fig. 2. This is because its launch of a second brand gives the incumbent more flexibility to adjust its premium product towards its monopoly quality \( q_{1t}^{*} \). And second, the entrant can be forced to reduce the quality of its product introduction. This follows from the fact that the second product has a lower quality than the premium product, that is, \( q_{3t} > q_{2t} \) and that the entrant’s best response is always 4/7 of the lowest quality offered in the market.
4. Extensions

I now extend the basic framework of Section 3 in three directions: First, I relax the assumption of costless imitation by the entrant and discuss how this affects equilibrium behavior. Second, I assume that the incumbent incurs positive costs when launching a second product and analyze the incumbent’s optimal response in presence of entry uncertainty. Third, I introduce the possibility of technological progress and analyze when it is optimal to introduce a fighter brand and simultaneously increase the quality of the premium brand.

4.1. Imitation costs and quality adjustments

In the basic framework I assumed that the entrant can costlessly imitate the incumbent’s quality. In a more general setup, suppose that the capability of the entrant to imitate is limited, for example, by the intensity of imitation, the size of spillovers, or the degree of intellectual property rights violation. Let \( \mu \) represent this degree of imitation, \( \mu \in [0,1] \), such that the higher the value of \( \mu \), the easier imitation is: For \( \mu = 0 \), no imitation is possible, whereas \( \mu = 1 \) corresponds to the case of full imitation. When choosing a quality \( q_i \), the entrant then has imitation costs

\[
c_i(q_i, q_e) = \begin{cases} 
\frac{1}{2} \gamma(q_i(1-\mu)q_e^2) & \text{if } q_e \leq q_i \\
\frac{1}{2} \gamma(q_i(q_e - \mu q_e^2)) & \text{if } q_e > q_i 
\end{cases}
\]

where \( \gamma \) is the entrant’s investment efficiency to develop a certain level of quality, see also Kovac and Zigic (2006) and Pepall (1997). Since the entrant usually is less efficient than the incumbent to produce a certain quality level, \( \gamma_e > \gamma \). The entrant’s profits in Period 2 then read as

\[
\pi_e = p_e q_e^2 - c_i(q_i, q_e).
\]

As in the basic model, it is never optimal for the entrant to produce a higher level of quality than the incumbent. This results from its lower level of efficiency. Moreover, when choosing the optimal product quality \( q_e \) the entrant now has to take into account the additional marginal imitation costs \( \gamma(q_i(1-\mu)q_e^2) \). The following proposition summarizes how equilibrium behavior will change.

**Proposition 6.** When it is costly for the entrant to imitate the quality of the incumbent:

1. The lower its entry quality will be,
2. The higher are the qualities of products offered by the incumbent independent of whether the incumbent reacts by price or by portfolio adjustment and,
3. The quality of the incumbent’s premium product increases towards its monopoly product, the higher are the entrant’s costs of imitation.

**Proof.** See Appendix A.

4.2. Launching costs and entry uncertainty

The discussion in Section 3.4 suggests that a portfolio adjustment is always better for the incumbent firm than a price response. Of course, this conclusion depends crucially on the costs that the incumbent has to bear when introducing a second product. If these costs are sufficiently low, the previous result still holds and the incumbent will always launch a new product to influence the entrant’s quality decision. However, if the launching costs are sufficiently high, such a portfolio adjustment will not be beneficial any more. Instead, the incumbent uses a price adjustment to influence the rival’s entry decision.

This argumentation suggests that only in case of high launching costs a price adjustment is better than a portfolio adjustment. This, however, neglects the fact, that the latter option of the incumbent is a long-run commitment to potential entry whereas a price adjustment is not. Indeed, the portfolio adjustment option is only beneficial for the incumbent if it is taken before entry – otherwise it would not influence the entrant’s quality choice any more. On the other hand, however, a price adjustment is a short-run decision in the sense that the incumbent reacts to the competition after entry actually occurred. To analyze this in more detail, suppose that the incumbent is uncertain about entry. This might happen, for example, in a situation in which the potential entrant has to bear positive entry costs. These costs might incur for advertising expenditures to inform consumers about the entrant’s product or for investments in transportation channels. In case the incumbent is uncertain about these entry costs, let \( \mu \in [0,1] \) be the priori probability that a potential competitor does not enter the market and \( (1-\mu) \) be the probability that entry occurs with certainty. Then the next proposition shows that the trade-off between a price and a portfolio adjustment depends crucially on the probability of entry as well as on the incumbent’s costs to launch a second product.

**Proposition 7.** When it is costly to launch a second product, the incumbent’s optimal reaction to entry is:

- **Price adjustment as in Proposition 2, if launching costs are sufficiently high.** \( F_L \geq F_L(\mu) \).
- **Portfolio adjustment as in Proposition 4, otherwise.**

**Proof.** See Appendix A.

Calculation shows that the critical value \( F_L(\mu) \) can be written as

\[
F_L(\mu) = \frac{5\theta(1-\beta)E^2}{4\gamma E^2} \gamma(1-\mu)(24 + \hat{\delta}(19 + 5\mu(1-\beta))).
\]

Then \( F_L(\mu) \) is increasing in the probability of entry. Hence, if the probability that the potential entrant actually enters the market is arbitrarily low, a price adjustment might be the better option even if the incumbent’s launching costs when introducing a second product are low. Of course, investing in such a situation in a product launch imposes costs which might be worthless in case the entrant actually does not enter. Here, instead of making such a risky long-term decision implying sunk costs \( F_L \), it might be more beneficial to wait until the entrant has actually entered and then use a price adjustment as a short-term tactical decision to defend the premium product.

Note that the incumbent never has an incentive to withdraw its second product even in case entry did not occur and independent of possible exit costs. The reason is obvious: Even with two products the incumbent can still earn monopoly profits. In fact, if the incumbent sets the price \( p_2^* \) such that the price-quality ratio of both products is identical, that is,

\[
p_2^* = p_2^* \frac{q_2^*}{q_1^*},
\]

no consumer will buy the low-quality product.

4.3. Development costs and technological progress

The discussion in Section 3.4 also suggests that an investment in quality improvement can never be a best response for the incumbent firm: If such an investment is profitable as a response to entry, it should have been done in the first period where the incumbent’s monopoly power gave maximal profits from such an investment.

This conclusion, of course, relies crucially on the assumption that the degree of the incumbent’s investments’ efficiency is constant over both periods. Suppose now that the degree of the efficiency increased from Period 1 to Period 2. Higher efficiency in product development could, for example, be the result of overall technological progress or firm-specific learning cost advantages. Let \( \lambda < 1 \) be the parameter that reflects
this increase in efficiency from $\gamma_t$ to $\lambda \gamma_t$. Using the previous results the incumbent then chooses its quality improvement to maximize second period profits for a given initial quality level $q_{t2}$.

$$\eta_{t2}^* = \frac{7}{48} \tilde{p} q_{t2} - \frac{1}{2} \lambda \gamma_t (\lambda q_{t2}^2 - q_{t2}^2).$$

The first-order condition implies that the incumbent optimally chooses $q_{t2}^* = \frac{7}{32} \tilde{p}$. Using its monopoly pricing in Period 1 the incumbent then introduces an initial quality level $q_{t2}$ to maximize overall profits

$$\eta_{t2} = \frac{1}{4} \tilde{p} q_{t2}^2 + \delta \frac{49}{4808 \lambda \gamma_t} \eta_{t2}^2 - \frac{1}{2} (1 - \delta) \gamma_t q_{t2}^2.$$

In the optimum, $q_{t2}^* = \frac{1}{32 \gamma_t \lambda} \tilde{p}^2$ which is lower than $q_{t2}^*$ as long as $\lambda \leq \lambda^* = \frac{7}{12} (1 - \delta)$.

**Proposition 8.** When the efficiency of product development increases, the incumbent’s optimal reaction to entry is:

- Quality adjustment of the premium product, if efficiency increase is sufficiently high, $(\lambda \leq \lambda^*)$.
- Portfolio adjustment as in Proposition 4, otherwise.

**Proof.** See Appendix A.

Hence, if the improvement in product development is sufficiently strong, an upgrade in product quality is beneficial for the incumbent. In this case, the price–quality ratio of the incumbent’s product increases compared to the case of portfolio adjustment whereas the price–quality ratio of the incumbent’s premium brand remains unchanged:

$$\frac{q_{t2}^*}{p_{t2}^*} = \frac{7}{8}, \quad \frac{q_{t2}^*}{p_{t2}^*} = \frac{1}{4} \tilde{p} q_{t2}^2, \quad \frac{7}{32} \tilde{p} > \frac{1}{8} \tilde{p}.$$

Of course, the incumbent then also has the possibility to simultaneously use a quality improvement of its premium brand and the introduction of a fighter brand. The next proposition shows that this is indeed the optimal combined response to entry even if the efficiency of product development is not that strong, that is $\lambda \leq \lambda^*$, with

$$\lambda^* = \frac{(12 - 5\delta)}{12} (1 - \delta) > \lambda.$$

**Proposition 9.** When the efficiency increase of product development is sufficiently high, $(\lambda \leq \lambda^*)$, and the incumbent can simultaneously react with several strategic options, a combined portfolio and quality adjustment is optimal.

As in Section 3.3, the incumbent uses its fighter brand as a firewall in Period 2. Due to the improvement in efficiency, however, the incumbent not only upgrades its premium product to $q_{tl}^*$ but also introduces its fighter brand on a higher quality level $q_{ll}^*$:

$$q_{ll}^* = \frac{(12 - 5\delta) \tilde{p}^2}{48 \lambda \gamma_t} > q_{tl}^* \quad \text{and} \quad q_{ll}^* = \beta q_{ll}^* > q_{ll}^*.$$

However, compared to the case in which the incumbent only chooses a portfolio adjustment without a quality improvement, the price–quality ratios of all three products remain unchanged

$$\frac{p_{tl}^*}{q_{tl}^*} = \frac{1}{2} \tilde{p} \left( 1 - \frac{\beta}{2} \right) \frac{q_{ll}^*}{q_{tl}^*} = \frac{1}{4} \tilde{p}^2 \frac{q_{ll}^*}{q_{tl}^*} = \frac{1}{8} \tilde{p}.$$

As a consequence, not only the incumbent earns higher profits from a quality improvement at the high end together with the introduction of a fighter brand, but that also the entrant is better off. This, of course, results from the fact, that the quality improvement of the premium product not only leads to a higher quality of the fighter brand, but, in turn, also improves the entrant’s product quality. Since marginal gains from selling a product are higher the higher its quality, the entrant benefits from the incumbent’s quality improvement as well.

**5. Conclusion**

Throughout the analysis I assumed that competition on the sales stage between the incumbent and the entrant is in prices and that both firms can adjust their price setting behavior in the short-run. How the nature of competition between firms might influence the results of this paper shows the following example, see Ritson (2009): In 2003 before the patent of its blockbuster drug Zocor in Germany expired, Merck decided to launch a second brand called Zocor MSD. To avoid too much cannibalization and to preserve customer loyalty with its premium product, Merck introduced its second product four months before the patent expiration of Zocor and priced it slightly underneath the one of the original premium brand. Once generics entered the market, the new product’s price dropped to 90% of Zocor’s. But this price cut was insufficient to seriously compete with the generics that invaded the market. More than 30 of these generic competitors, accustomed to competing on price, divide almost the entire generic market among themselves. Merck’s desire to protect its premium brand failed. Even more important, when Merck realized that it had set the wrong initial price, it was incapable of quick course correction. This example suggests that the mode of competition on the sales stage – price or quantity – as well as the mode of competition on the differentiation stage – vertical or horizontal – is crucial for the success of a second product launch to fight against low-price competition. In particular, the trade-off between a price or portfolio adjustment might crucially depend on the underlying mode of competition.

Another restrictive assumption of the previous analysis is the fact that consumers are totally aware of the differences in product qualities. In particular, I assumed that consumers know that the incumbent’s second product is simply a product of lower quality at a lower price than the high quality product. The following story, however, shows what happens if this assumption is not satisfied, see also Ritson (2009): In 1994 Kodak launched a second film to fend off its best seller Gold Plus film against its Japanese rival Fuji with its Fujicolor Super G film. This new film called Funtime had a lower quality compared to its own premium brand since Kodak manufactured Funtime with an older, less effective technology than Gold Plus. Moreover, it was sold at the same price as Fuji’s offering. Although, in principle, Funtime was created explicitly to win back customers that had switched to Fuji’s low-price alternative, two years later, in 1996, Kodak withdrew Funtime from the market. The reason was that most consumers were unaware of the quality differences between the two products of Kodak: they simply saw Funtime as a high quality Kodak film at a lower price. This seriously damaged Kodak’s reputation for high quality and Gold Plus sales were more damaged than Fuji’s. This example shows that the success of a portfolio adjustment depends crucially on how the incumbent firm brands its products. As in the case of Kodak, the incumbent can use umbrella branding so that the premium product as well as the second, lower quality product are labeled with a single brand name. Although umbrella branding might play an informational role in markets in which consumers are uncertain about product characteristics, see e.g. Cabral (2009), it leads to increased cannibalization of the premium.
brand in the context of the present model. But then it might be beneficial for the incumbent to sell its two products under different names or, if this is not possible, to use price adjustment to protect its premium product from low-price competitors.

I also assumed in the previous analysis that customers follow the standard model of rational choice by evaluating prices and qualities of different products in a rational way. Of course, under bounded rationality different behavioral patterns are important for consumers’ choice, e.g. the compromise effect, see Geyskens, Gielens, and Gijsbrechts (2010), and may determine the success of the incumbent’s strategic options as well the entrant’s entry strategy.

Appendix A

Proof of Proposition 1. Suppose the incumbent offers only one product ($q_M, p_M$). In the absence of entry, the incumbent then chooses a price $p_M$ to maximize per period gross profits

$$\pi_M = p_M \left( \bar{\theta} - \frac{p_M}{q_M} \right),$$

for a given quality level $q_M$. Hence

$$p_M = \frac{1}{2} \pi_M \text{ and } q_M = \frac{1}{2} \pi.$$

Using this monopoly price the optimal level of quality then maximizes

$$\pi_M = p_M x_M(1 + \theta) - c_1(q_M) = \frac{1}{4} \bar{\theta}^2 (1 + \theta) q_M - \frac{1}{2} \gamma q_M^2.\tag{1}$$

The first-order condition reads as

$$\frac{\partial}{\partial q_M} \pi_M^1 = \frac{1}{4} \bar{\theta}^2 (1 + \theta) - \gamma q_M = 0.$$

hence

$$q_M = \frac{\bar{\theta}^2 (1 + \theta)}{4 \gamma}, p_M = \frac{\bar{\theta}^2 (1 + \theta)}{8 \gamma}$$

and monopoly profits are

$$\pi_M^* = \frac{\theta^2 (1 + \theta)^2}{32 \gamma}.$$ 

To see the second part of the proposition, it is sufficient to show that the incumbent will never introduce two products. So suppose for the question per period then leads to the following two propositions.

Proof of Proposition 2. To solve the game use backward induction.

Stage 2c Given qualities $q_1 > q_{i1}$ and prices $p_{12}$, and $p_{12}',$ demand in Period 2 is

$$x_{12} = \bar{\theta} - \frac{p_1^2}{q_1 - q_{i1}}, x_{12}' = \frac{p_{12}^2 - p_{12}'^2}{q_1 - q_{i1}} - \frac{p_{12} - p_{12}'}{q_1 - q_{i1}}.$$ 

Stage 2b Given qualities $q_i > q_{i1}$, the entrant maximizes profits

$$\pi_{i1} = p_{i1}^2 q_{i1} x_{i1}' = p_{i1}^2 \left( \frac{p_{12}^2 - p_{12}'^2}{q_1 - q_{i1}} - \frac{p_{12} - p_{12}'}{q_1 - q_{i1}} \right)$$

with respect to its product price $p_{12}'$. This leads to the first-order condition

$$\frac{\partial \pi_{i1}}{\partial p_{12}'} = \frac{p_{12}^2 - 2 p_{12}' q_{i1}}{q_1 - q_{i1}} = 0,$$

hence, $p_{12}' = \frac{2 p_{12} q_{i1}}{q_1 - q_{i1}}.$ Similarly, the incumbent maximizes its second period profits

$$\pi_{12}^* = p_{12}^2 q_{i1} x_{i1} = p_{12}^2 \left( \frac{p_{12}^2 - p_{12}'^2}{q_1 - q_{i1}} \right)$$

with respect to its product price $p_{12}'. The first-order condition then reads as

$$\frac{\partial \pi_{12}}{\partial p_{12}'} = \frac{p_{12}^2 q_{i1} - q_{i1}^2}{q_1 - q_{i1}} = 0,$$

hence, $p_{12}' = \frac{1}{2} p_{12} + \frac{1}{2} \bar{\theta} (q_1 - q_{i1}).$ Hence, in a price equilibrium

$$p_{12}' = \frac{2 p_{12} q_{i1}}{q_1 - q_{i1}}, p_{12}^* = \frac{\partial \pi_{12}}{\partial p_{12}'} = \frac{p_{12}^2 (q_1 - q_{i1})}{4 q_1 - q_{i1}}.$$ 

Stage 2a Given qualities $q_i$, the entrant chooses $q_{i1}$ to maximize

$$\pi_{i1} = (q_1 - q_{i1}) (q_1 - q_{i1}) (q_1 - q_{i1})$$

and optimal monopoly prices can be computed as

$$p_{12} = \frac{1}{2} \pi_M, p_{12}' = \frac{1}{2} \pi_M.$$

Note that the second-order conditions $\frac{\partial^2 \pi_{i1}}{\partial p_{12}'^2} = - \frac{2}{(q_1 - q_{i1})} < 0$ and $\frac{\partial^2 \pi_{i1}}{\partial q_{i1}^2} = -2 \frac{1}{(q_1 - q_{i1})^2} < 0$ are both satisfied. This solution, however, implies zero demand for the low quality product, $x_{i1} = 0$ since $\theta_{i1} = \theta_{iD}$. Moreover, the demand for the high quality product is $x_{iH} = \bar{\theta}/2$, that is, the incumbent will not introduce a second product of lower quality.

Q.E.D. 

Stage 1a Using the results from Proposition 1 for Period 1 the incumbent then maximizes total profits from monopoly and competition over both periods

$$\pi_1 = \frac{1}{4} \bar{\theta} q_1 + \frac{\bar{\theta}^2}{4 q_1} q_1 - \frac{1}{2} \gamma q_1^2.$$
taking into account the entrant’s best response \( q_1^* \). The optimal product quality \( q_1^* \) then satisfies the first-order condition
\[
\frac{\partial}{\partial q_1} \pi_{11} = \frac{1}{4} \pi^2 + \frac{\partial^2}{48} \gamma q_1 = 0
\]

hence
\[
q_1^* = \frac{\beta^2 (12 + 7\delta)}{48\gamma_1}.
\]

Inserting this into the entrant’s best quality response gives
\[
q_1^* = \frac{\beta^2 (12 + 7\delta)}{48\gamma_1}.
\]

Equilibrium prices then are
\[
p_1^* = \frac{\beta^2 (12 + 7\delta)}{96\gamma_1}, \quad p_2^* = \frac{\beta^2 (12 + 7\delta)}{192\gamma_1}, \quad p_2^* = \frac{\beta^2 (12 + 7\delta)}{672\gamma_1},
\]
resulting in equilibrium profits
\[
\pi_{11}^* = \frac{\beta^2 (12 + 7\delta)^2}{4608\gamma_1}, \quad \pi_{11}^* = \frac{\beta^2 (12 + 7\delta)^2}{2304\gamma_1}.
\]

Q.E.D.

**Proof of Proposition 4.** In solving the game, I distinguish two cases: either \( q_{11}^* > q_{11} - q_{12} \) (Case 1) or \( q_{11}^* > q_{12} - q_{11} \) (Case 2):

Stage 2c/b — Case 1 Given \( q_{11}^* > q_{12} > q_{11} \), and prices \( p_{1h}, p_{2h}^*, \) and \( p_{1h}^* \), the incumbent maximizes profits
\[
\pi_{1h} = p_{1h}^* q_{1h} - q_{1h} p_{1h}^* - q_{1h} q_{1h}^* - \frac{\beta^2}{32} \left( p_{1h}^* - p_{2h}^* \right) \left( q_{1h} - q_{1h}^* \right)
\]

with respect to its product price \( p_{2h}^* \). This leads to the first-order condition
\[
\frac{\partial \pi_{1h}}{\partial p_{2h}^*} = 2p_{2h}^* q_{1h} - q_{1h} p_{2h}^* = 0,
\]

hence \( p_{2h}^* = \frac{1}{2} p_{1h} q_{1h} \). Similarly, the incumbent maximizes its overall profits in Period 2,
\[
\pi_{2h} = p_{1h}^* q_{1h} - q_{1h} p_{1h}^* - \frac{\beta^2}{32} \left( p_{1h}^* - p_{2h}^* \right) \left( q_{1h} - q_{1h}^* \right)
\]

with respect to its prices \( p_{1h}^* \) and \( p_{2h}^* \). The first-order conditions then are
\[
\frac{\partial \pi_{2h}}{\partial p_{1h}^*} = \frac{\beta^2}{32} \left( q_{1h} - q_{1h}^* \right) - \frac{p_{1h}^* - p_{2h}^*}{48\gamma_1} = 0,
\]
\[
\frac{\partial \pi_{2h}}{\partial p_{2h}^*} = 2p_{2h}^* (q_{1h} - q_{1h}^*) - \frac{p_{1h}^* - p_{2h}^*}{48\gamma_1} (q_{1h} - q_{1h}^*) = 0.
\]

The best response functions are obtained from these conditions and read as:
\[
p_{1h}^* = \frac{p_{2h}^*}{2} + \frac{\beta}{2} (q_{1h} - q_{1h}^*),
\]
\[
p_{2h}^* = \frac{p_{2h}^*}{2} (q_{1h} - q_{1h}^*) + \frac{p_{1h}^*}{2} (q_{1h} - q_{1h}^*).
\]

Solving for \( p_{2h}^* \) and \( p_{1h}^* \), then gives equilibrium prices
\[
p_{1h}^* = \frac{1}{2} p_{1h}^* + \frac{\beta}{2} (q_{1h} - q_{1h}^*) + \frac{\beta^2}{4} (q_{1h} - q_{1h}^*),
\]
\[
p_{2h}^* = \frac{2\beta}{32} (q_{1h} - q_{1h}^*) (q_{1h} - q_{1h}^*) + \frac{\beta^2}{4} (q_{1h} - q_{1h}^*).
\]

Of course, \( p_{1h}^* > p_{2h}^* \) since \( (q_{1h} - q_{1h}^*) (q_{1h} - q_{1h}^*) + \beta^2 (q_{1h} - q_{1h}^*) \) > 0. Substituting these equilibrium prices into the demand function gives the equilibrium demand
\[
q_{1h} = \frac{2}{\beta^2} \left( q_{1h} - q_{1h}^* \right), \quad q_{1h} = \frac{2}{\beta^2} \left( q_{1h} - q_{1h}^* \right).
\]

Stage 2c/b — Case 2 Given \( q_{1h} > q_{1h} > q_{1h} \) and prices \( p_{1h}, p_{2h}^*, \) and \( p_{1h}^* \), demand now is
\[
q_{1h} = p_{1h} (q_{1h} - q_{1h}^*) + \frac{1}{2} p_{1h} (q_{1h} - q_{1h}^*) - 2p_{2h}^* (q_{1h} - q_{1h}^*) (q_{1h} - q_{1h}^*)
\]
\[
\frac{\partial \pi_{2h}}{\partial p_{1h}^*} = p_{1h}^* (q_{1h} - q_{1h}^*) + \frac{1}{2} p_{1h}^* (q_{1h} - q_{1h}^*) - 2p_{2h}^* (q_{1h} - q_{1h}^*) (q_{1h} - q_{1h}^*) = 0.
\]

The entrant’s profit function
\[
\pi_{2h} = p_{1h}^* q_{1h} - q_{1h} p_{1h}^* - \frac{\beta^2}{32} \left( p_{1h}^* - p_{2h}^* \right) \left( q_{1h} - q_{1h}^* \right)
\]

is maximized with respect to its product price \( p_{2h}^* \) if the first-order condition is satisfied:
\[
\frac{\partial \pi_{2h}}{\partial p_{2h}^*} = p_{2h}^* (q_{1h} - q_{1h}^*) + \frac{1}{2} p_{1h}^* (q_{1h} - q_{1h}^*) - 2p_{2h}^* (q_{1h} - q_{1h}^*) (q_{1h} - q_{1h}^*) = 0.
\]

The best response function of the entrant then is
\[
p_{2h}^* = \frac{1}{2} p_{1h}^* (q_{1h} - q_{1h}^*) + \frac{1}{2} p_{1h}^* (q_{1h} - q_{1h}^*) - 2p_{2h}^* (q_{1h} - q_{1h}^*) (q_{1h} - q_{1h}^*) = 0.
\]

The incumbent’s maximization of profits
\[
\pi_{2h} = p_{1h}^* q_{1h} - q_{1h} p_{1h}^* - \frac{\beta^2}{32} \left( p_{1h}^* - p_{2h}^* \right) \left( q_{1h} - q_{1h}^* \right)
\]
\[
\pi_{2h} = p_{1h}^* q_{1h} - q_{1h} p_{1h}^* - \frac{\beta^2}{32} \left( p_{1h}^* - p_{2h}^* \right) \left( q_{1h} - q_{1h}^* \right)
\]

with respect to its prices \( p_{1h}^* \) and \( p_{2h}^* \). The first-order conditions then are
\[
\frac{\partial \pi_{2h}}{\partial p_{1h}^*} = p_{1h}^* (q_{1h} - q_{1h}^*) + \frac{1}{2} p_{1h}^* (q_{1h} - q_{1h}^*) + 2p_{2h}^* (q_{1h} - q_{1h}^*) \left( q_{1h} - q_{1h}^* \right) = 0,
\]
\[
\frac{\partial \pi_{2h}}{\partial p_{2h}^*} = p_{1h}^* (q_{1h} - q_{1h}^*) + \frac{1}{2} p_{1h}^* (q_{1h} - q_{1h}^*) + 2p_{2h}^* (q_{1h} - q_{1h}^*) \left( q_{1h} - q_{1h}^* \right) = 0.
\]
which imply \( p_{EL}^2 = \frac{1}{2} p_{EL}^3 + \frac{1}{2} (q_{EL} - q_{EL}) \) and \( p_{EL}^3 = \frac{1}{2} p_{EL}^3 q_{EL} \). Solving all three conditions for equilibrium prices yields

\[
\begin{align*}
q_{EL}^2 &= \frac{\overline{p}_{EL}^2 \left( q_{EL} - q_{EL} \right) + q_{EL} \left( q_{EL} - q_{EL} \right) + 3 \overline{p}_{EL}^3 \left( q_{EL} - q_{EL} \right)}{2 \overline{q}_{EL} \left( q_{EL} - q_{EL} \right) + q_{EL} \left( q_{EL} - q_{EL} \right) + 2 \overline{p}_{EL}^3 \left( q_{EL} - q_{EL} \right)} \\
q_{EL}^3 &= \frac{\overline{p}_{EL}^3 \left( q_{EL} - q_{EL} \right) + q_{EL} \left( q_{EL} - q_{EL} \right) + 3 \overline{p}_{EL}^3 \left( q_{EL} - q_{EL} \right)}{2 \overline{q}_{EL} \left( q_{EL} - q_{EL} \right) + q_{EL} \left( q_{EL} - q_{EL} \right) + 2 \overline{p}_{EL}^3 \left( q_{EL} - q_{EL} \right)} \\
q_{EL}^4 &= \frac{\overline{p}_{EL}^4 \left( q_{EL} - q_{EL} \right) + q_{EL} \left( q_{EL} - q_{EL} \right) + 3 \overline{p}_{EL}^3 \left( q_{EL} - q_{EL} \right)}{2 \overline{q}_{EL} \left( q_{EL} - q_{EL} \right) + q_{EL} \left( q_{EL} - q_{EL} \right) + 2 \overline{p}_{EL}^3 \left( q_{EL} - q_{EL} \right)}
\end{align*}
\]

Stage 2a. Given \( q_{HI} > q_{EL} \) the entrant can either choose \( q_{HL} = q_{EL} \) with \( q_{HL} < q_{EL} \) or \( q_{HL} = q_{EL} \) with \( q_{HL} = q_{HL} \). If the entrant chooses \( q_{EL} < q_{EL} \), then using the equilibrium prices above, the first-order condition for profit maximization gives

\[
\frac{\partial \pi_{EL}}{\partial q_{EL}} = \frac{\partial \left( \overline{p}_{EL}^2 \left( q_{EL} - q_{EL} \right) \right)}{\partial q_{EL}} = \overline{p}_{EL}^2 \left( 4 q_{EL} - 2 q_{EL} \right) = \overline{p}_{EL}^2 \left( q_{EL} - q_{EL} \right) = 0
\]

and it optimally chooses \( q_{EL}^* = \frac{1}{2} q_{EL} \). Hence,

\[
\pi_{EL} = \frac{1}{48} \overline{p}_{EL}^2 q_{EL}. \quad (C1)
\]

If, on the other hand, \( q_{EL} \in \left[ q_{EL}, q_{HI} \right] \) then using the equilibrium prices above, the entrant’s profits read as

\[
\pi_{EL} = \frac{\overline{p}_{EL}^2 q_{EL} \left( q_{EL} - q_{EL} \right) + q_{EL} \left( q_{EL} - q_{EL} \right) + 3 \overline{p}_{EL}^3 \left( q_{EL} - q_{EL} \right)}{2 \overline{q}_{EL} \left( q_{EL} - q_{EL} \right) + q_{EL} \left( q_{EL} - q_{EL} \right) + 2 \overline{p}_{EL}^3 \left( q_{EL} - q_{EL} \right)}.
\]

In this case, the first-order condition is

\[
\frac{\partial \pi_{EL}}{\partial q_{EL}} = \frac{\partial \overline{p}_{EL} \left( q_{EL} - q_{EL} \right)}{\partial q_{EL}} = \overline{p}_{EL} \left( q_{EL} - q_{EL} \right)
\]

with \( A = -7 \overline{q}_{EL}^2 q_{HI} + 5 q_{EL}^3 q_{EL} - 4 q_{EL}^3 q_{HI} + 4 \overline{p}_{EL}^3 q_{HI} \). Note that for \( q_{HI} > q_{EL} \) it makes sense for \( q_{HI} > q_{EL} \) and for \( q_{HI} < q_{EL} \), it follows \( A > 3 q_{HI} (q_{HI} - q_{EL})^2 > 0 \) for continuity then requires that the entrant optimally chooses \( q_{HI} \) such that \( A = 0 \), that is,

\[
(3 q_{EL}^2 q_{HI} - 4 q_{EL}^3 q_{HI})^2 - 2 (2 q_{EL}^3 - 4 q_{EL}^2 + 2 q_{EL}) (q_{HI} q_{EL} + 3 q_{EL} q_{HI})
\]

\[
= (2 q_{EL} q_{HI} - 4 q_{EL} q_{HI} + q_{HI} q_{EL} + (q_{EL}^3))^2
\]

Note that \( q_{HI} < q_{EL} \) requires \( q_{HI} = q_{EL} \) and \( q_{HI} = q_{EL} \), which is always satisfied in the optimal. Then the entrant chooses \( q_{EL} = q_{EL} \), if \( q_{HI} > q_{EL} \) and otherwise \( q_{EL} = q_{EL} \).

Stage 1c. Note that \( q_{EL} \) is monotonely increasing in \( q_{HI} \).

\[
\frac{\partial q_{EL}}{\partial q_{HI}} = \overline{p}_{EL} \left( q_{EL} - q_{EL} \right) > 0
\]

with \( q_{EL} = 0 \) for \( q_{HI} = q_{HI} \) and \( q_{EL} = \frac{1}{2} \overline{p}_{EL} \) for \( q_{HI} \). Let \( q_{EL} \) be the critical value given by \( q_{EL} \) or \( q_{EL} \), that is,

\[
q_{EL} = \frac{5 (\overline{q}_{EL})^2 + q_{EL} \left( q_{EL} - q_{EL} \right) + 3 (q_{EL} - q_{EL})}{4 (q_{EL} - q_{EL})^2 + 6 q_{EL} q_{EL} - 4 q_{EL}}
\]

with \( X = 25 (\overline{q}_{EL})^2 + 34 q_{EL} q_{EL} - 23 q_{EL} q_{EL} \). Note that in equilibrium \( X > 0 \) since otherwise \( C2 \) would not have a solution. Moreover note that \( q_{EL} \in (0, q_{EL}) \), in \( q_{EL} \) and \( q_{EL} \) is always strictly greater than \( X > 0 \). To see the last claim note that for \( q_{HI} > q_{EL} \) we have \( q_{HI} \leq q_{EL} \) if \( 0 < 12 (2 q_{HI} - q_{HI}) (2 q_{HI} + q_{HI}) \) and for \( q_{HI} < q_{EL} \) we have \( q_{HI} \geq q_{EL} \) if \( 0 < 12 (2 q_{HI} - q_{HI}) (2 q_{HI} + q_{HI}) \), which is true in both cases. Hence, \( q_{HI} = q_{HI} \). Substituting this \( q_{HI} \) in condition \( C2 \) and rearranging terms then yields

\[
q_{HI} \left( q_{HI} - q_{HI} \right)^2 \left( 3 q_{HI} q_{HI} + 2 \left( q_{HI} \right)^2 - 2 \overline{p}_{EL}^2 \right) \]

\[
= \left( 73 q_{HI} - 192 q_{HI} \right)^2 + 57 q_{HI} q_{HI} + 2 q_{HI} q_{HI} \right)^2 + 89 q_{HI} q_{HI}\)

\]

Hence, the optimal \( q_{HI} \) is linear in \( q_{HI} \) and solution of

\[
73 - 192 q_{HI} + 57 \alpha^2 + 89 \alpha^3 = 0.
\]

The only solution with \( \alpha > 4/7 \) then \( \alpha = 0.77738 \). The optimal \( q_{HI} \) then is

\[
q_{HI} = 4 \alpha^2 + 5 \alpha^3 - 3 \alpha - \alpha (1 - \alpha) \sqrt{25 \alpha^2 + 34 \alpha - 23} + 4 \alpha^2 + 6 \alpha - 4
\]

with \( b_{HI} = 0.548056 q_{HI} \).

The incumbent then positions its second product with an optimal fighter brand quality slightly above \( q_{HI} \) such that the entrant’s best response
is \( q^*_{3t} = \frac{4}{3} q^*_{M} = \frac{4}{3} \beta q^*_{Mt} \). The relationships between marginal consumers then is

\[
\bar{\nu}_M = \frac{1}{4} \bar{\nu}_L (2 + \beta) > \bar{\nu}_E = \frac{1}{12} \bar{\nu}_L (7 \beta - 2) > \bar{\nu}_I = \frac{1}{9} \bar{\nu}_L
\]

which follow from \( \beta \in (\frac{1}{2}, 1) \). Moreover, the resulting profits are

\[
\pi_{E3} = \frac{1}{48} \bar{\nu}^2 q_{3t}, \quad \pi_{I3}^2 = \frac{1}{48} \bar{\nu}^2 (12 - 5\beta) q_{3t}
\]

in case of a price adjustment and

\[
q_{3t} = \alpha q_{Mt}, \quad q^*_{3t} = \alpha q^*_{Mt}, \quad q_{E3} = \frac{4}{7} \alpha q^*_{Mt} \quad \text{and} \quad \pi_{E3}^2 = (2 - \beta) \nu^2_q \cdot \frac{\alpha^2}{2} \pi_M^2, \quad \pi_{I3}^2 = \frac{\alpha^2}{2} \pi_M^2
\]

in case of a portfolio adjustment.

Q.E.D.

Proof of Proposition 6. Consider, for example, the case of price adjustment, Option 1. The first-order condition for \( q_{E1} \) reads as

\[
\bar{\nu}^2 q_{E1}^2 \left( q_{E1} - q_{E1} \right)^2 \gamma_q (1 - \mu) = 0
\]

Using the envelope theorem it then follows that

\[
\frac{\partial q_{E1}^*}{\partial \gamma_q} = \frac{- \partial q_{E1}^* (1 - \mu)}{\partial \mu} = - \frac{q_{E1}^* (1 - \mu) (4 q_{E1} - q_{E1})^3}{72 \gamma_q^2 q_{E1} + 4 \gamma_q (1 - \mu) (4 q_{E1} - q_{E1}) (4 q_{E1} - q_{E1})^3} < 0
\]

Taking this into account, the incumbent maximizes

\[
\pi_{E1} = \frac{1}{4} \bar{\nu}^2 q_{E1} + \bar{\nu}^2 \gamma_q^2 q_{E1} - \frac{1}{2} \gamma_q (1 - \mu) q_{E1}^2
\]

Since

\[
\frac{\partial \pi_{E1}}{\partial q_{E1}} > 0 \quad \text{and} \quad \frac{\partial \pi_{E1}}{\partial \mu} < 0
\]

it follows that

\[
\frac{\partial q_{E1}^*}{\partial \gamma_q} > 0 \quad \text{and} \quad \frac{\partial q_{E1}^*}{\partial \mu} < 0
\]

Moreover,

\[
\frac{\partial \pi_{E1}^2}{\partial q_{E1}} = \frac{\partial \pi_{E1}^2}{\partial \mu} = \frac{6 \bar{\nu}^2 q_{E1}^2 (4 q_{E1} - q_{E1})^3}{(4 q_{E1} - q_{E1})^4} \left( q_{E1} \frac{\partial q_{E1}^*}{\partial \gamma_q} - q_{E1} \frac{\partial q_{E1}^*}{\partial \mu} \right) > 0
\]

Q.E.D.

Proof of Proposition 7. The incumbent firm has two strategic options: either not to introduce a second product and use a price adjustment (Case 1) or to launch a firhter brand and use a portfolio adjustment (Case 2). Let \( \mu \in [0, 1] \) be the a priori probability that the rival firm will never enter the market and \( (1 - \mu) \) be the probability that behavior that entry occurs with certainty.

Case 1 Using the results of Proposition 1 in case the incipient remains monopolist with probability \( \mu \) and offers only its brand product \( q_{E1} \), and using the results of Proposition 2 in case the entrant enters with probability \( (1 - \mu) \) and the market becomes a duopoly with two products \( (q_{E1}, q_{I1}) \), expected equilibrium profits are

\[
\pi_{E1}^2 (q_{E1}) = \frac{1}{4} \bar{\nu}^2 q_{E1} + (1 - \mu) \bar{\nu}^2 \gamma_q^2 q_{E1} - \frac{1}{2} \gamma_q (1 - \mu) q_{E1}^2
\]
where it is already assumed that the entrant in case of entry chooses \( q_1^e, q_2^e = \frac{\lambda}{2} q_{1L} \). Then the optimal quality of the premium product is given by \( \frac{\partial}{\partial q_{1H}} E_{1L} = 0 \), hence,

\[
q_{1H}^* = \frac{\tilde{\theta}^2}{4\gamma I_1} \left(1 + \delta - \frac{5}{12} (1-\mu)\delta \right)
\]

and equilibrium expected profits read as

\[
E_{1H}^* = \frac{\tilde{\theta}^4}{4608 \gamma_1} (12\delta - 5(1-\mu)\delta + 12)^2.
\]

**Case 2** Using the results of **Proposition 1** in case the incumbent remains monopolist with two products (\( q_{1L}, q_{2L} \)) with probability \( \mu \), and using the results of **Proposition 4** in case the entrant enters with probability \( (1- \mu) \) and the market becomes a duopoly with three products (\( q_{1H}, q_{2H}, q_{1L} \)), expected equilibrium profits for the incumbent are

\[
E_{1L}(q_{1L}, q_{1H}, q_{2L}) = \frac{\tilde{\theta}^2}{4} (1 + \delta) q_{1L} \left(1 - \mu\right) \left(1 + \delta - \frac{5}{12} (1-\mu)\delta \right) q_{1H} - \frac{1}{2} \gamma q_{1L} - F_L,
\]

where the price equilibrium and the best responses \( q_{1H}^*(q_{1L}) = \frac{\lambda}{2} q_{1L} \) and \( q_{2L}^*(q_{1L}) = \beta q_{1L} \) are already included. The first-order condition for profit maximization \( \frac{\partial}{\partial q_{1H}} E_{1L} = 0 \) then yields

\[
q_{1H}^* = \frac{\tilde{\theta}^2}{4\gamma I_1} \left(1 + \delta - \frac{5}{12} (1-\mu)\delta \right)
\]

and equilibrium expected profits read as

\[
E_{1L}^* (q_{1L}) = \frac{\tilde{\theta}^4}{4608 \gamma_1} (12\delta - 5\delta(1-\mu)\delta + 12)^2 - F_L.
\]

Comparison with Case 1 then shows that \( E_{1L}^* (q_{1L}) \geq E_{1L}^* (q_{1L}) \) whenever

\[
F_L \leq \frac{5\delta(1-\beta)\tilde{\theta}^2}{4608 \gamma_1} (1-\mu)(24 + \delta(19 + 5\mu(1-\beta))) = \tilde{F}_L.
\]

Note that the critical value such that the incumbent uses a price adjustment is decreasing in \( \mu \) since

\[
\frac{\partial}{\partial \mu} \left(1 - \mu \right) (24 + \delta(19 + 5\mu(1-\beta))) < 0.
\]

Q.E.D.

**Proof of Proposition 8.** To prove this proposition it is sufficient to show that for \( \lambda < 7(1-\delta)/12 \) a quality adjustment actually leads to higher profits of the incumbent firm. In case of a quality improvement overall profits are

\[
n_{12}^* = \frac{1}{32} \lambda \tilde{\theta}^4 \left( \frac{49}{144} \delta + \frac{\lambda}{1-\delta} \right)
\]

whereas in case of a portfolio adjustment profits are

\[
n_{13}^* = \frac{\tilde{\theta}^4}{4608 \gamma_1} (12 + \delta(12-5\beta))^2,
\]

see **Proposition 4**. Then \( n_{12}^* > n_{12}^* \) if

\[
\lambda \left( - (12-5\beta)((12-5\beta)\delta + 24) + ((12-5\beta)\delta + 12)^2 \right) + 49(1-\delta) > 0.
\]

The left hand side of this inequality is monotone in \( \lambda \) and positive for \( \lambda = 0 \). Moreover, the left hand side is positive for \( \lambda = 7(1-\delta)/12 \) since

\[
\frac{7}{12} \left( - (12-5\beta)((12-5\beta)\delta + 24) + ((12-5\beta)\delta + 12)^2 \right) + 49
\]

\[
= (\delta + 1.5484)(\delta + 0.043439) > 0.
\]

Q.E.D.

**Proof of Proposition 9.** Using the proof of **Proposition 4**, Stage 1c, the incumbent positions its optimal fighter brand quality such that \( \bar{q}_{1L} = \beta q_{1L} \), whereas the entrant’s best response is \( q_{2L} = \frac{\lambda}{2} q_{1L} \). The resulting profits are

\[
n_{123}^* = \frac{1}{48} \tilde{\theta}^2 \left( \beta q_{1L}^2 \right)^2. \quad \hspace{1cm} n_{123}^* = \frac{(12-5\beta)^2}{48} \tilde{\theta} q_{1L}^2 - \frac{1}{2} \gamma \left( \lambda q_{1H}^2 - q_{123}^2 \right).
\]

Maximizing \( n_{123}^* \) with respect to \( q_{1L}^* \) then yields

\[
q_{123}^* = \frac{1}{48} \tilde{\theta} q_{123}^2 + \beta q_{1L}^2 \tilde{\theta}^4 - \frac{1}{2} (1-\delta) \gamma (q_{123})^2. \quad \hspace{1cm}
\]

These profits are maximized when the incumbent chooses

\[
q_{123}^* = \frac{1}{48} \tilde{\theta} q_{123}^2 + \beta q_{1L}^2 \tilde{\theta}^4 - \frac{1}{2} (1-\delta) \gamma (q_{123})^2.
\]

which is lower than the quality adjustment in Period 2, \( q_{123} < q_{1L}^* \), if

\[
\lambda < \frac{12-5\beta}{12} \hspace{1cm} (1-\delta).
\]

Equilibrium prices can then be derived from the result of Stage 2c/b — Case 1 of **Proposition 4** as

\[
p_{1H}^* = \frac{(12-5\beta)(2-\beta)\tilde{\theta}^2}{192 \gamma_1}, \quad \hspace{1cm} p_{1L}^* = \frac{(12-5\beta)\tilde{\theta}^2}{192 \gamma_1}
\]

\[
and \hspace{1cm} p_{123}^* = \frac{(12-5\beta)\tilde{\theta}^4}{672 \gamma_1}.
\]

Equilibrium profits then are

\[
n_{123}^* = \frac{1}{32} \lambda \tilde{\theta}^4 \left( \frac{(12-5\beta)^2}{144} \delta + \frac{\lambda}{(1-\delta)} \right) \quad \hspace{1cm} n_{123}^* = \frac{(12-5\beta)\tilde{\theta}^4}{2304 \gamma_1}.
\]

To show that \( n_{123}^* > n_{12} \) holds for all \( \lambda \geq \lambda^* \) it is sufficient to assume \( \lambda = \lambda^* \) since \( n_{123}^* = n_{123}^*(\lambda) \) is decreasing in \( \lambda \). Calculation shows that \( n_{123}^*(\lambda^*) > n_{12}^* \) is equivalent to

\[
144 - 12(1-2\delta)(12-5\beta) + (1-\delta)\delta(12-5\beta)^2 > 0.
\]

Since the left hand side of this inequality is increasing in \( \beta \), and for \( \beta = 0 \) always satisfied since

\[
1 > 1 - \delta^2 - \delta,
\]

it follows that \( n_{123}^* > n_{12}^* \). Moreover,

\[
n_{123}^* > n_{12}^* = \frac{1}{2304 \gamma_1} (12 + \delta(12-5\beta))
\]

iff

\[
\lambda < \frac{12-5\beta}{12 + \delta(12-5\beta)}
\]
which is always satisfied since \( \lambda \leq \frac{12 - 5\beta}{12 - \delta} + \frac{5\beta}{12 + (12 - 5\beta)\delta} \) and

\[
\frac{(12 - 5\beta)}{12} (1 - \delta) < \frac{(12 - 5\beta)}{12 + (12 - 5\beta)\delta}. 
\]

iff \( \tilde{q} > q^* \) is also equivalent to \( \lambda \leq \frac{12 - 5\beta}{12 + (12 - 5\beta)\delta} \).

Q.E.D.

References


Full Length Article

A multi-category customer base analysis

Chang Hee Park a,*, Young-Hoon Park b,1, David A. Schweidel c,2

a School of Management, The State University of New York at Binghamton, P.O. Box 6000, Binghamton, NY 13902, United States
b Samuel Curtis Johnson Graduate School of Management, Cornell University, 361 Sage Hall, Ithaca, NY 14853, United States
c Goizueta Business School, Emory University, GA 30322, United States

A R T I C L E   I N F O

Article history:
First received in April 16, 2013 and was under review for 5 months
Available online 12 March 2014

Area Editor: Olivier Toubia

Keywords:
Customer base analysis
Timing models
Multi-category choice models
Latent space models
Bayesian estimation

A B S T R A C T

Customer base analysis is an essential tool to measure and develop relationships with customers. While various models have been proposed in a noncontractual setting, they focus primarily on analyzing transactional patterns associated with a single product category or a firm-level activity, such as the times at which purchases are made at a particular retailer. This research proposes a modeling framework for customer base analysis in a multi-category context. Specifically, we model the time between a customer’s purchases at the firm and the product categories that comprise her shopping basket arising from multi-category choice decisions. The proposed model uses a latent space approach that parsimoniously captures the dynamics of multi-category shopping behavior due to the interplay between purchase timing and shopping basket composition. We also account for interdependence among multiple categories, temporal dependence across category choices, and latent customer attrition. Using category-level transaction data, we show that the proposed model offers excellent fit and performance in predicting customer purchase patterns across multiple categories. The forecasts and inferences afforded by our model can assist managers in tailoring marketing efforts across categories.

Copyright © 2014 Elsevier B.V. All rights reserved.

1. Introduction

Customer base analysis is an essential tool to manage and develop relationships with customers by predicting their future purchase patterns (Schmittlein & Peterson, 1994). The uses of the analyses range from aggregate-level sales trajectories to individual-level customer valuation (Fader & Hardie, 2009). The “buy ‘til you die” framework is a widely adopted modeling approach for customer base analysis in a noncontractual setting. This class of models is based on the assumption that customers repeatedly make transactions at a firm while they are in an active state (alive), and at some unobserved time point they become permanently inactive (dead). Though many variants of “buy ‘til you die” models have been proposed (e.g., Fader, Hardie, & Lee, 2005; Fader, Hardie, & Shang, 2010; Schmittlein, Morrison, & Colombo, 1987), they primarily focus on analyzing transactional patterns associated with a single product category or a firm-level activity, such as the times at which purchases are made at a particular retailer.

Oftentimes, however, customers’ shopping behavior involves the purchase of multiple product categories at a single transaction. In such cases, marketers can gain a more comprehensive understanding of its customer base by analyzing customers’ purchase patterns within individual categories. The inferences would be of particular importance to firms making marketing and operational decisions at the category level. For example, the analyses could be used for customer targeting and scoring (e.g., Reinartz & Kumar, 2003; Schweidel & Knox, 2013) on a category-by-category basis. If a customer is predicted to be more actively purchasing a specific category than others, sending the customer marketing offers relevant to the category could more effectively facilitate the realization of the future transactions within this category. Such category-level marketing efforts could be further improved by recognizing the relationships that may exist across categories, both during the current shopping trip (i.e., what categories tend to be co-purchased) and across shopping trips (i.e., are category purchases on the previous trip informative of category purchases on the current trip). For example, cross-category dependencies may suggest promoting some product categories to customers to increase their likelihood of making purchases in multiple categories, thereby increasing their total expenditures and value to the retailer. The objective of this research is therefore to propose a modeling framework for multi-category customer base analysis in a noncontractual setting.

Modeling customer purchase patterns across categories requires a number of careful considerations. First, a customer’s arrival process to the firm and her category purchase decisions are likely to be independent. Consider a customer whose shopping trips to the store are...
driven by multiple categories. Depending on the nature of the relationship among the categories for the customer, the combination of product categories in her shopping basket may be informative of how long it will be until her next shopping trip to a store. For example, if a customer were to purchase beef and pork from a butcher to prepare meals, the time until the customer returns to the butcher to make another purchase may be longer than if she were to just have purchased one of the meats on her previous trip. Preparing a meal with one of the meats may delay the consumption of the other, resulting in a longer time to consume both meats before making the next shopping trip. A customer may return to a store sooner following a trip in which she bought several items together if those items are jointly consumed, such as ingredients going into a salad. In turn, the customers’ interarrival time to the store may affect their category purchase decisions upon a visit. Second, in a multi-category context, a customer’s purchase decisions on a given shopping trip may be correlated across categories. Accounting for such associations is therefore important to capture the simultaneous choice of categories in a shopping basket. Third, repeatedly observed shopping behavior from the same customer could be correlated over time, as a customer’s current choice decisions are likely to be associated with past outcomes both within and across categories. Finally, latent customer attrition needs to be considered in a noncontractual setting.

In this research, we develop a model for multi-category customer base analysis which accommodates the aforementioned complications by generalizing the “buy ‘til you die” framework to a multi-category context. Toward this end, we jointly model customers’ interarrival process to the firm, multi-category purchase decisions, and latent attrition while explicitly accounting for the shopping dynamics arising from the interplay of purchase timing and incidence across categories. The key challenge in modeling the interarrival process and its association with multi-category purchase behavior is that the number of possible compositions of shopping baskets increases exponentially as the number of categories increases, and so does the complexity of the model. To address this issue parsimoniously, we employ a latent space approach that can be generalized to a large number of product categories with ease. In modeling the shopping basket composition, we account for independent category choices within a transaction and the correlation of repeatedly observed outcomes over time both within and across categories. We also allow customers’ category purchase decisions to depend on the timing of their past purchase of the category. Therefore, our model captures the sequential interaction between shopping basket choice and interarrival times.

We apply our model to category-level customer data from a leading beauty care company in Korea. Though the firm’s product categories do not have clear demand relationships with each other, we observe systematic variation in customers’ interarrival times related to the combination of categories purchased at their prior transaction. Through a series of benchmark models, we assess the need to incorporate the association of shopping basket choice and interarrival times as well as cross-category dependence both within and across transactions. We find that ignoring the effect of shopping basket composition on the arrival process impairs the prediction of firm-level transactional patterns. Moreover, omitting the cross-category dependence in category choice decisions results in the overestimation of the purchases of shopping baskets containing only one category and the underestimation of the purchases of shopping baskets consisting of multiple categories. Our proposed model offers superior fit and performance in predicting individual customers’ purchase patterns of shopping baskets, in addition to informing our understanding of the relationships that exist among multiple aspects of customer behavior in a multi-category transactional context. The results can be used by marketers for the customization of targeting and communication strategies. For example, firms can identify those customers most likely to make purchases within each of the categories, leveraging their purchase histories across all categories, thereby allowing marketers to tailor their communications based on the categories in which a customer is likely to purchase activity.

The remainder of this article is organized as follows. Section 2 gives an overview of prior literature related to our research. In Section 3, we describe the data used in our empirical analysis. Section 4 provides a detailed specification of our model. In Section 5, we illustrate the performance of the model and discuss inferences afforded by investigating purchase patterns of shopping baskets. Finally, Section 6 concludes and suggests future research directions.

2. Related work

The current research draws on prior work that has examined customers’ purchase patterns and multi-category behavior. We briefly review relevant literature on both streams of research and discuss the contribution of our work relative to them.

Predicting customers’ purchase patterns from a timing perspective has long been of interest to marketing researchers. Much research in this domain has relied on the assumption of an exponential interarrival process due to its parsimony and performance (e.g., Gupta, 1991; Schmittlein, Bemmar, & Morrison, 1985), which underlies the negative binomial distribution (NBD) model (e.g., Ehrenberg, 1988; Morrison & Schmittlein, 1988). Several researchers have relaxed the assumption of a memory less timing process, such as by taking steps to allow for nonstationarity in the customer arrival process (e.g., Fader, Hardie, & Huang, 2004; Moe & Fader, 2004; Schweidel & Fader, 2009), incorporating time-varying explanatory variables (e.g., Gupta, 1988, 1991), and considering alternative baseline timing processes (e.g., Allenby, Leone, & Jen, 1999; Jain & Vilcassim, 1991; Seetharaman & Chintagunta, 2003). An important extension of timing models for customer base analysis in transactional settings was the addition of latent attrition, resulting in “buy ‘til you die” models such as the Pareto/NBD (Schmittlein et al., 1987) and the BG/NBD (Fader et al., 2005). These models have served as the foundation for analyzing transactional patterns within a single product category or a firm-level activity in a noncontractual setting (e.g., Abe, 2009; Batislam, Denizel, & Filiztekin, 2007; Ma & Büschnken, 2011; Singh, Borle, & Jain, 2009; Van Oest & Knox, 2011; Wübben & Wangenheim, 2008).

Subsequent research has generalized the univariate models of customers’ interarrival times to a multivariate context. Prior research has used multivariate distributions, such as the Farlie–Gumbel–Morgenstern distributions (e.g., Chintagunta & Haldar, 1998) and the Sarmanov distributions (e.g., Park & Fader, 2004; Schweidel, Fader, & Bradlow, 2008) to correlate the timing processes of different categories or transactional activities. Though these models reveal the association between the timing processes, perhaps owing to the model complexity that would accompany generalizing the multivariate distributions beyond two activities, such research has largely been restricted to bivariate applications. Moreover, they did not take into account customer attrition and thus have limitations for customer base analysis. Our model, using a latent space approach (e.g., Bradlow & Schmittlein, 2000; DeSarbo & Wu, 2001; Li, Sun, & Wilcox, 2005), generalizes to a multivariate context in a parsimonious manner. This is valuable from a managerial perspective, as it allows firms to assess the relationships that may exist across a broader range of product categories and assess their impact on customers’ multi-category purchase patterns.

In addition to generalizing univariate multi-event timing models, our work is related to the research on multi-category choice behavior (e.g., Fader & Lodish, 1990; Narasimhan, Neslin, & Sen, 1996). Recognizing that a customer’s purchase decisions across categories are not independent, multivariate choice models have incorporated heterogeneity and coincidence in purchase decisions (e.g., Chib, Seetharaman, & Strijnev, 2002; Manchanda, Ansari, & Gupta, 1999; Moon & Russell, 2003). Throughout the paper, we use the terms, “the composition of shopping baskets” and “shopping basket composition” to refer to the combination of product categories purchased in the same shopping basket on a shopping trip.
Our research differs from extant multi-category choice models in two notable ways. First, rather than focusing solely on the incidence decisions, we jointly model the shopping basket composition and the interarrival process, allowing for their interactive effect on purchase patterns across categories. Though extant research has considered both multi-category choices and the arrival process (e.g., Chintagunta, 1999; Gupta, 1988; Wagner & Taudes, 1986), prior work has not explored the relationship between shopping basket composition and the interarrival process. Second, we incorporate latent attrition in customers’ relationship with the firm to accommodate the importance of accounting for the latent attrition in forecasting future transactional activities (Fader & Hardie, 2009). In doing so, our modeling framework enables us to forecast both firm- and category-level metrics of managerial interest.

3. Data

Our data come from a leading beauty care company in Korea. The data set consists of the transaction histories of 2,870 customers who purchased beauty care products of a high-end brand designed for women at department stores in Korea from January 2008 to December 2010. For our purpose, we categorized 79 cosmetic products (i.e., SKU) purchased by the customers during the data period into the following four categories: (1) basics, (2) creams, (3) serums, and (4) makeups. This categorization is also used on the firm’s dashboard for marketing planning and performance evaluation. Basics consist of lotions and toners which are frequently purchased together as must-have skin care items for women. Creams and serums are more advanced and topical (e.g., anti-aging and anti-wrinkle) skin care products, and makeups contain products for facial makeup and cleansing.

Table 1 summarizes the descriptive statistics of our data. On average, customers made 9.06 transactions with the firm and purchased 1.64 different categories per transaction. Out of the total transactions by the customers, 54% were the purchase of only one category, 30% involved two categories, 13% involved three categories, and the remaining 3% were the purchase of all four categories. Among the categories, creams were most frequently purchased (4.62 times per consumer) and makeups were least purchased (2.46 times per consumer). Looking at the combinations of categories purchased together on a shopping trip, we find that customers purchased 5.10 different compositions of shopping baskets through their 9.06 transactions over the data period.

To explore the customers’ shopping patterns, we examine the relationship between the number of transactions and the average number of categories purchased per transaction across the customers in Fig. 1. We find a significant negative correlation (−0.29) between the number of transactions and the number of categories purchased per transaction. This indicates that while some customers purchase more categories on fewer transactions, others shop more frequently and buy less per transaction, illustrating the need to simultaneously consider the frequency of visits (i.e., the interarrival process) and shopping basket composition.

Given customers’ purchases of multiple categories per transaction (an average of 1.64 different categories per transaction), for each shopping basket, in Table 2 we present the frequency with which it is purchased, the number of customers purchasing it and the time taken until the next transaction since they purchased it. The first column of the table lists all possible basket compositions denoted by a vector whose four elements indicate the purchase of basics, creams, serums, and makeups in the order, respectively. For example, [1,0,0,0] in the table denotes a shopping basket containing basics only. The table shows that the number of observations varies considerably across shopping baskets, ranging from 430 for [1,0,1,1] to 4,489 for [0,1,0,0]. The most frequently purchased shopping basket ([0,1,0,0]) contains creams only and was purchased by 62% of the customers, while only 12% of customers purchased the shopping basket [1,0,1,1]. Furthermore, the time taken until the next transaction, conditional on a transaction of a specific shopping basket, ranges from 66.3 days to 95.3 days. Though we generally observe longer times until the next transaction following the purchase of shopping baskets with more categories, we find a variation in the time until the next transaction following the purchase of shopping baskets containing the same number of categories. As customers purchased multiple different compositions of shopping baskets over the data period (an average of 5.10 different compositions of shopping baskets) and the interarrival times vary considerably with the combination of categories purchased, this suggests the need to consider the association between basket composition and the interarrival process in forecasting customers’ future purchase patterns across categories.

4. Model development

Our proposed model consists of the following three components: (1) a customer’s arrival process to a firm, (2) latent attrition, and (3) her choice decisions across multiple categories conditional on arrival. We first present the specification of the model components and discuss how they interact with each other. We then describe our computational approach to estimating the model.

![Fig. 1. Scatter plot between the no. of transactions and the mean no. of categories purchased per transaction.](image-url)
4.1. Timing model

During the period \((0, T]\), where 0 corresponds to the beginning of the model calibration period and \(T\) is the censoring point that corresponds to the end of the data period, we observe customer \(i\) who makes \(J_i\) transactions at a firm at times \(t_1, t_2, \ldots, t_{J_i}\). In each transaction at the firm, the customer makes purchase decisions across \(K\) product categories. We define a binary variable \(C_{ijk}\) to indicate whether or not customer \(i\) purchases category \(k\) at her \(j\)th transaction. Customer \(i\)'s shopping basket at the \(j\)th transaction can be represented by a vector of category choice outcomes, \(C_i = \{C_{i1}, C_{i2}, \ldots, C_{iK}\}\), where \(C_{ijk}\) equals 1 if customer \(i\) purchases category \(k\) at her \(j\)th transaction and 0 otherwise. It is important to note that, in our offline shopping context, we have \(\sum_{k=1}^{K} C_{ijk} \geq 1\) for any transaction, because a customer’s shopping trip to the firm is observed only when she purchases at least one category.\(^4\)

We first model customer \(i\)'s arrival process to the firm by specifying the timing behavior for a set of \(J_i−1\) interarrival times, \(t_2−t_1, t_3−t_2, \ldots, t_{j_i}−t_{j_i−1}\), and the right-censored observation \(T−t_{j_i}\). We assume that customer \(i\)'s interarrival time between her \(j\)th and \((j + 1)\)th transactions follows an exponential distribution with arrival rate \(\lambda_j\). Then, the density function for the interarrival times and the survival function for the right-censored observations are given by:

\[
\begin{align*}
    f(t_{j_i+1}−t_j; \lambda_j) &= \lambda_j \exp\left(-\lambda_j(t_{j_i+1}−t_j)\right) \\
    S(T_{j_i+1}; \lambda_j) &= \exp\left(-\lambda_j(T−t_{j_i})\right).
\end{align*}
\]

When a customer repeatedly shops across multiple product categories, her shopping basket composition at a transaction may be related to the time until her next shop. To take into account this effect, we specify \(\lambda_j\) as:

\[
\lambda_j = \lambda_j f(C_{ij}) \exp(X_j\alpha),
\]

where \(\lambda_j\) is customer \(i\)'s baseline arrival rate, \(f(\cdot)\) is a function of the prior shopping basket choice \((C_{ij})\), \(X_j\) is a vector of covariates which may affect the purchase timing behavior, and \(\alpha\) is a vector of the corresponding parameters.

At the heart of our specification of \(\lambda_j\) in Eq. (2) is the function \(f(C_{ij})\) which captures the association between the customer’s shopping basket choice and the time until her next shopping trip. The function \(f(C_{ij})\), as well as the covariates \(X_j\), allow for nonstationality in the timing process and for a customer’s subsequent decisions to be informed by her past decisions. We next discuss alternative specifications of \(f(C_{ij})\).

4.1.1. An additive model

A simple approach that incorporates the effects of previously purchased categories into the arrival rate \(\lambda_j\) would be to include \(C_{ijk}\)’s as additive terms with exponential base in \(f(C_{ij})\) (to ensure \(\lambda_j > 0\)). Formally,

\[
f(C_{ij}) = \exp\left(a_1 C_{i1} + a_2 C_{i2} + \ldots + a_K C_{iK}\right),
\]

where \(a_k\) is a coefficient which captures the effect of customer \(i\)'s purchase of category \(k\) on her next arrival rate.

This specification accounts for the “main effects” of each purchased category on the timing process. If we were to restrict \(a_k = a\) for all \(k\), this would be equivalent to counting up the number of categories in the shopping basket. However, Eq. (3) does not take into account the interplay that may exist among multiple categories, thereby ignoring the possible joint effects of categories on the interarrival process. To account for this, one could include all possible interaction terms of categories purchased, or equivalently assume different arrival rates following the purchase of each combination of categories. While this approach allows one to account for the dependence of the arrival rate on the composition of the previous shopping basket, the possible shopping basket compositions (and hence model parameters) increase exponentially with the number of categories. To address this issue parsimoniously, we next propose a modeling approach that places product categories in a latent space (e.g., Bradlow & Schmittlein, 2000; Li et al., 2005) and use this to parameterize their relationship to the interarrival process.

4.1.2. A latent space model

We model the effect of shopping basket composition on the arrival rate using the Euclidean distances between the categories and the origin in the space. To formalize this, we specify \(f(C_{ij})\) as:

\[
f(C_{ij}) = \left|\left|\sum_{k=1}^{K} P_k C_{ijk}\right|\right|^{-1},
\]

where \(P_k\) is the position of category \(k\) in an \(n\)-dimensional latent space, represented by \(n\) coordinates to be estimated, and \(\left|\left|\cdot\right|\right|\) denotes the Euclidean norm of a point (i.e., the Euclidean distance between the point and the origin).

We illustrate the intuition behind Eq. (4) with an example of two categories on a two-dimensional space: \(P_1\) and \(P_2\) in Fig. 2 represent categories 1 and 2, respectively, on the space where the positions of the categories are denoted by their \(x\)- and \(y\)-coordinates. The evaluation of the Euclidean norm in Eq. (4) and the inverse relationship between the arrival rate and interarrival time indicate that when a customer purchases category \(k\) only, the effect of the purchase on the interarrival time until the next transaction is given by \(\left|\left|P_k\right|\right| = (x_k^2 + y_k^2)^{1/2}\). Thus, a larger (smaller) value of \(\left|\left|P_k\right|\right|\) implies that the customer’s interarrival time after purchasing category \(k\) tends to be longer (shorter).

Next, when a customer purchases both categories 1 and 2, its effect on the interarrival time is given by the Euclidean norm of the summed vector, \(\left|\left|P_1 + P_2\right|\right| = ((x_1 + x_2)^2 + (y_1 + y_2)^2)^{1/2}\). Note that the Euclidean norm varies depending on the angle, \(\delta\), between the two vectors, \(P_1\) and \(P_2\). For a given length of the vectors, as \(\delta\) decreases, \(\left|\left|P_1 + P_2\right|\right|\) increases and thus the interarrival time after purchasing both categories becomes larger, which we would anticipate for product categories in which consuming one may delay the consumption of the other. In contrast, larger values of \(\delta\) would reflect shorter expected interarrival times, which would occur when consuming one category is expected to accelerate the consumption of another. Thus, \(\delta\) reflects
the combined effect of the categories on the arrival rate. The same logic would apply to cases with more than two categories.

Note that when a one-dimensional space is employed, each category is represented by a single coordinate (i.e., x-coordinate) on the space. Hence, similar to the case in Eq. (3), the specification of \( f(C_i) \) in Eq. (4) only accounts for the main effects of the categories using the same set of covariates with a different functional form (i.e., \( f(C_i) = \frac{1}{\sum_{k=1}^{K} x_k C_{ik}^{-1}} \)). The latent space can be extended to more than one dimension to provide increased flexibility. The selection of the number of dimensions in the latent space would depend on the goal of the analysis. A two-dimensional space would allow for a relatively simple representation of categories in the space. However, a higher dimensional space may provide better model fit as the number of categories increases. For identification, the number of dimensions cannot exceed the number of categories considered.

The Euclidean norms of a set of vectors are invariant under rotation and reflection of the space. To identify the model, we impose the following restrictions. For a two-dimensional space, we restrict the x-coordinate of category 1 to be positive and the y-coordinate of the category to zero. This takes into account the rotation of the space. We also restrict the y-coordinate of category 2 to be positive so as to account for the reflection of the space over the y-axis. Finally, because \( \lambda_i \) is proportional to the product of \( \lambda_i \) and \( \prod_{k=1}^{K} x_k C_{ik}^{-1} \), there are an infinite number of combinations of \( \lambda_i \) and \( x_k \) that give the same value of \( \lambda_i \). To consider this, we set the x-coordinate of category 1 to one. The remaining coordinates in the latent space are treated as model parameters to be inferred from the data. Similar identification conditions can be derived for a general case with an n-dimensional space as follows. First, we restrict the first coordinate of category 1 to one and all other coordinates of the categories to zero. For category k such that \( 1 < k \leq n \), we restrict the \( k \)-th coordinate of the category to be positive and the \( m \)-th coordinate to be zero for all \( m \) such that \( m > k \). No identification condition is required on the coordinate of category \( k > n \). Using a latent space model allows us to parameterize the effects of shopping basket composition parsimoniously as \( n \) increases.

4.2. Customer attrition

As our empirical context is transactional (i.e., noncontractual) in nature, customers may cease shopping at the firm and silently defect at an unobserved time. The failure to incorporate latent attrition can result in biased parameter estimates for the arrival process. We therefore incorporate latent attrition by assuming that customer \( i \) becomes permanently inactive after a transaction with probability \( r_i \), consistent with the individual-level attrition process of the BG/NBD model (Fader et al., 2005)7:

\[
P(\text{Customer } i \text{ drops out after her } j^{\text{th}} \text{ transaction}) = r_i(1-r_i)^{j-1}.
\]  

Because the conditional logit model, which has been employed in a few marketing studies to account for interdependent choices in multiple categories (e.g., Moon & Russell, 2008; Russell & Petersen, 2000). The model captures cross-category dependence by incorporating the choice outcomes in other categories into the choice decision in the focal category. The model also allows us to derive the closed form expression for the unconditional joint distribution of a customer’s category choices, which greatly facilitates our objective of predicting future shopping basket composition.\(^8\)

We specify the probability that customer \( i \) purchases category \( k \) at the \( j \)-th transaction conditional on her choices of other categories, using the following logit form:

\[
Pr(C_{ijk} = 1|C_{ij-1} \neq k) = \left[ 1 + \exp\left( -\pi_{ik} + \sum_{k' \neq k} \theta_{ik'} C_{ij} + \sum_{k' \neq k} \psi_{ik'} C_{ij-1} \right) \right]^{-1},
\]  

where \( \pi_{ik} \) captures the time-varying category-specific effect for category \( k \) at customer \( i \)’s \( j \)-th transaction and \( \sum_{k' \neq k} \theta_{ik'} C_{ij} \) reflects the effects of the choices of other categories on the choice decision in category \( k \). Hence, \( \theta_{ik} \) measures the degree of interdependence between categories \( k \) and \( k' \). A positive \( \theta_{ik} \) implies that the purchase of category \( k' \) increases the probability of purchasing category \( k \) at the same transaction. A negative \( \theta_{ik} \) reflects a negative association between the choices of categories \( k \) and \( k' \). Finally, \( \sum_{k' \neq k} \psi_{ik'} C_{ij-1} \) captures the effects of the choices of other categories in the prior transaction on the choice of category \( k \) in the current transaction. A positive (negative) \( \psi_{ik} \) implies that the purchase of category \( k' \) in the prior transaction increases (decreases) the probability of purchasing category \( k \) in the current transaction.

To consider the effect of time-varying covariates on the category choice probability, we further specify the category-specific intercept \( \pi_{ik} \) as:

\[
\pi_{ik} = \beta_{ik0} + Z_{ik}\beta_k.
\]

where \( \beta_{ik0} \) is a customer-specific baseline purchase tendency for category \( k \), \( Z_{ik} \) is a vector of time-varying covariates, and \( \beta_k \) is a vector of the corresponding parameters for category \( k \).

Because the conditional probabilities in Eq. (6) are mutually dependent across categories, we cannot use the equation to predict a customer’s choice decisions across multiple categories. By assuming \( \theta_{ik} = \theta_{k} \) and applying the theorem by Besag (1974), however, it can be shown that the unconditional joint distribution of \( C_{ij} \) is given by:

\[
P\left( C_{ij} = j, \sum_{k=1}^{K} C_{ik} \geq 1 \right) = \exp\left( \frac{\pi_j C_j + \frac{1}{2} C_j^{1/2} \Theta C_j + C_{j-1} \Psi C_{j-1}}{\sum_{C_{j-1}(0,0,...,0)} \exp(\pi_j C_j + \frac{1}{2} C_j^{1/2} \Theta C_j + C_{j-1} \Psi C_{j-1})} \right).
\]
where

\[
\pi_y = \begin{bmatrix} \pi_{y1} \\ \vdots \\ \pi_{yK} \end{bmatrix}, \quad \theta = \begin{bmatrix} 0 & \theta_{12} & \cdots & \theta_{1K} \\ \theta_{12} & 0 & \cdots & \theta_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{1K} & \theta_{2K} & \cdots & 0 \end{bmatrix}
\]

and

\[
\Psi = \begin{bmatrix} 0 & \psi_{12} & \cdots & -\psi_{1K} \\ \psi_{21} & 0 & \cdots & -\psi_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{K1} & \psi_{K2} & \cdots & 0 \end{bmatrix}
\]

and the summation in the denominator is across the possible shopping baskets constructed from \(K\) categories excluding the “no-purchase” case. The derivation of Eq. (8) is provided in Appendix A.

While the proposed multi-category choice model in Eq. (8) allows us to flexibly capture the interdependence across category choices within a transaction and their temporal dependence across transactions, one potential problem is that the number of parameters in both matrices \(\Theta\) and \(\Psi\) quadratically increases as the number of categories increases. To address this issue, we parameterize \(\theta_{kk}\) and \(\psi_{kk}\) as:

\[
\theta_{kk} = \xi_{kk} \psi_{kk} + \eta_{kk} + \xi_{k}\|P_k + P_k\| \quad \text{and}
\]

\[
\psi_{kk} = \xi_{kk} + \xi_{k} + \xi_{k} \|P_k + P_k\|.
\]

where \(\xi_{kk}, \xi_{k}\) and \(\xi_{k}\) are category-specific parameters, \(\|P_k + P_k\|\) is the Euclidean norm of the summed vector, \(P_k + P_k\), and \(\xi_{k}\) and \(\xi_{k}\) are coefficients for \(\|P_k + P_k\|\).

Our parameterization of \(\theta_{kk}\) and \(\psi_{kk}\) can be thought of as a decomposition into main effects associated with categories \(k\) and \(k'\), and the interaction between categories \(k\) and \(k'\) which is accounted for by the Euclidean norm \(\|P_k + P_k\|\). \(\theta_{kk}\) is comprised of the fixed effects of categories \(k\) and \(k'\) (\(\xi_{kk}, \xi_{k}\) and \(\xi_{k}\)) and the effect of their spatial relationship in the latent space employed in the timing model (\(\xi_{k}\|P_k + P_k\|\)). Similarly, \(\psi_{kk}\) is decomposed into the fixed effects of categories \(k\) and \(k'\) (\(\xi_{kk}, \xi_{k}\) and \(\xi_{k}\)) and the effect of their spatial relationship in the latent space used in the timing model (\(\xi_{kk}\|P_k + P_k\|\)). Note that the specification of the category-fixed effects for \(\psi_{kk}\) accommodates \(K\) more parameters than the specification for \(\theta_{kk}\), because \(\psi_{kk}\) is asymmetric with respect to \(k\) and \(k'\) and \(\theta_{kk}\) is symmetric with respect to \(k\) and \(k'\). We set \(\xi_{k1}\) to zero for identification which serves as a baseline across different categories. No identification condition is needed for \(\theta_{kk}\) given its symmetry. The Euclidean norm, \(\|P_k + P_k\|\), is included to allow for the possibility that the relationship between categories \(k\) and \(k'\) in the multi-category choice process is correlated with the relationship between the categories in driving interarrival times.\(^9\)

4.4. Customer heterogeneity and correlation structure

To incorporate customer heterogeneity into our model, we specify the customer-specific model parameters as follows. For the customer-specific baseline arrival rate in Eq. (2), we assume that \(\lambda_i\) follows a lognormal distribution to ensure that \(\lambda_i\) is positive. Second, we reparameterize the dropout probability \(r_i\) in Eq. (5) as \(r_i = \exp(-\exp(\omega_i))\) to ensure \(r_i \in [0,1]\), and assume that \(\omega_i\) follows a normal distribution.

\(^9\) In parameterizing \(\theta_{kk}\) and \(\psi_{kk}\), one may consider employing new latent spaces intended to capture the interactive effects of categories \(k\) and \(k'\) on the parameters, instead of the one used in the timing model. We have chosen not to pursue this direction because it significantly increases the model complexity and the proposed specification in Eq. (10) allows us to track customers’ purchase patterns across categories accurately.

Third, we assume that the customer-specific intercepts \(\beta_{0i1}, \beta_{0i2}, \ldots, \beta_{0iK}\) in Eq. (7) follow a normal distribution. Taken together, we assume

\[
\begin{bmatrix} \log \lambda_i \\ \omega_i \\ \beta_{0i1} \\ \vdots \\ \beta_{0iK} \end{bmatrix} \sim \text{MVN} \left( \begin{bmatrix} \mu_\lambda \\ \mu_\omega \\ \mu_{\beta_{0i1}} \\ \vdots \\ \mu_{\beta_{0iK}} \end{bmatrix}, \Sigma \right)
\]

to account for the correlation of the model parameters. Incorporating the covariance structure allows us to capture the interdependence of timing (arrival and attrition) and category choice decisions within a customer.

Similar to the matrices \(\Theta\) and \(\Psi\), the number of variance–covariance parameters in the matrix \(\Sigma\) quadratically increases as the number of categories considered increases. When this causes significant computational inefficiency in model estimation, one can consider the reparameterization of the covariance between \(\beta_{0ik}\) and \(\beta_{0ik'}\) using the category-specific parameters, in a similar manner used in Eq. (9). Alternatively, one may employ a block diagonal covariance structure in the matrix. For example, one may assume that covariance parameters between \(\beta_{0ik}\) and \(\beta_{0ik'}\) are zero at the expense of not incorporating the correlation across category choices within individuals.

4.5. Computational approach

We adopt a Bayesian approach and use the Markov chain Monte Carlo (MCMC) methods to estimate the proposed model. The model is estimated using the Gibbs sampler (Gelfand & Smith, 1990) with the Metropolis–Hastings steps (Hastings, 1970) in the publicly available software WinBUGS. The samples obtained from the MCMC algorithm are then used to compute summary measures of the parameter estimates. The results reflect the output of MCMC draws for 30,000 iterations, with a burn-in period of 30,000 iterations. To complete the Bayesian specification of the model, we assume noninformative conjugate priors to the parameters. For aggregate-level parameters and mean parameters, we use a diffuse normal density prior. For the variance–covariance matrix, we assume that the inverse of the matrix follows a Wishart distribution.

5. An empirical application

In this section, we provide an empirical demonstration of the proposed model, using data described in Section 3. We use the first 30-month period of the data for model calibration and the remaining six-month period for model validation. The calibration data consist of 2,621 customers who made repeated transactions during the calibration period.

5.1. Covariates

We begin by describing vectors of covariates, \(X_i\) and \(Z_{ii}\) included in Eqs. (2) and (7), respectively. \(X_i\) contains the lagged interarrival time (i.e., customer \(i\)'s interarrival time between her \((j-1)\)th and \(j\)th transactions), denoted by \(\text{LIT}_i\), to take into account the dependence of the current interarrival time on the past observation of the outcome variable. \(Z_{ii}\) contains (1) the lagged choice of category \(k\) (i.e., \(C_{ij-1,k}\)) to consider the temporal dependence of choices within a category, (2) the elapsed time between customer \(i\)'s \(j\)th transaction and her last purchase of category \(k\) before the transaction \(\text{ET}_{ik}\) to consider the relationship between the frequency of purchasing category \(k\) and the category choice probability in a given transaction and (3) the quadratic term of \(\text{ET}_{ik}\) to capture the possible nonlinear time effect of the prior category purchase on the current purchase decision. For example, the probability of purchasing the category at a given transaction may increase as the time since the last purchase of the category increases.
On the other hand, a very long elapsed time may indicate that the customer has stopped purchasing the category, resulting in a small purchase probability. By definition, ET_{ijk} = LIT if customer i purchased category k at her \((j-1)\)th transaction and ET_{ijk} = LIT otherwise. If other time-varying covariates were available and believed to affect the interarrival or multi-category choice processes, they could be incorporated into \(X_0\) and \(Z_{0k}\) (e.g., Schweidel & Knox, 2013).

Incorporating the aforementioned covariates into the model allows us to parsimoniously capture the association of customers’ purchase timing and category choice decisions across transactions, and the correlation of repeatedly observed outcome measures. Because all covariates are constructed based on the outcome variables we model in Section 4, the values of the covariates can be simulated beyond the calibration period to predict the customers’ future purchase patterns.

5.2. Model comparison

We fit a series of models to assess the importance of accounting for the interplay between shopping basket composition and interarrival times, the cross-category effect on category choices within a transaction, and the temporal dependence across category choices. The alternative models we examine therefore mainly vary with respect to the following three aspects: (1) whether or not (and how) to incorporate the effect of shopping basket composition on the arrival rate, (2) whether or not to account for the interdependence across category choices within a transaction, and (3) whether or not to consider the temporal dependence across category choices.

The first benchmark model (Model 1) is constructed using the BG/NBD model proposed by Fader et al. (2005). We employ four independent BG/NBD models for the prediction of purchase patterns of the four categories. Model 1 thus completely ignores all of the three aforementioned sources of correlation as well as the effects of covariates included in \(X_0\) in Eq. (2) and \(Z_{0k}\) in Eq. (7). Model 2 is the simplest version of our model which also fails to consider the three modeling aspects and the effects of the covariates. This model is formulated by setting \(f(C_{ij}) = 1\) and \(X_{ij} = 0\) in Eq. (2), \(\theta_{ijk} = 0\) and \(\psi_{ijk} = 0\) in Eq. (6), and \(Z_{ijk} = 0\) in Eq. (7). Note that this model differs from Model 1 in that the model components are still related at the margin due to the correlation that may exist among the underlying model parameters, as specified in Eq. (10). Model 3 extends Model 2 by simply including the covariate vectors \(X_{ij}\) and \(Z_{ijk}\) in Eqs. (2) and (7), respectively.

Model 4 extends Model 3 by considering the temporal dependence across category choices through \(\psi_{ijk}\) s, but still ignores the impact of shopping basket choices on interarrival times \(f(C_{ij}) = 1\) and the cross-category effects within a transaction \((\theta_{ijk} = 0)\). Model 5 extends Model 4 by adopting our multi-category choice model to account for the interdependence across category choices through \(\theta_{ijk}\) s, but fails to consider the association of shopping basket composition with interarrival times \((f(C_{ij}) = 1)\). In contrast to Model 5, Model 6 incorporates the effect of shopping basket choices on interarrival times by employing the proposed latent space approach, but ignores both the cross-category effects within a transaction \((\theta_{ijk} = 0)\) and the temporal dependence across category choices \((\psi_{ijk} = 0)\).

Model 7 allows for the cross-category effects within a transaction as well as the temporal dependence across category choices. However, when it comes to the effect of category choices on the arrival rate, the model only considers the main effects of purchased categories and does not account for their interplay by assuming that \(f(C_{ij})\) is given as in Eq. (3). Model 8 is the most generalized version constructed using our modeling framework. The model explicitly considers the effect of shopping basket composition on the arrival rate by including all two-, three- and four-way interaction terms of \(C_{ij}\) s as well as their main terms within the exponential term in Eq. (3). The model also explicitly considers both the cross-category effects within a transaction and the temporal dependence across category choices by employing \(\theta_{ijk}\) and \(\psi_{ijk}\) in Eq. (6) without the parameterization specified in Eq. (9). As a result, the model is more heavily parameterized than our proposed model and the number of parameters exponentially increases as the number of categories increases. Finally, Model 9 is our proposed model in which the latent space approach is used in specifying \(f(C_{ij})\) and the parameterization of \(\theta_{ijk}\) and \(\psi_{ijk}\) as shown in Eq. (9).

To provide a comparison of the performance of the benchmark models, we compute the log marginal density of the models. A larger value of the log marginal density indicates a better model fit. We also compute the mean absolute error (MAE = |Predicted − Observed|) with respect to the number of purchases of each shopping basket by individual customers, averaged across shopping baskets and customers.\(^\text{11}\) For the computation of the MAE of Model 1, which consists of four independent category choice processes, we assume that categories are purchased together in the same shopping basket if the purchases of the categories are predicted to occur in the same week.

Table 3 reports the model comparison results. We find that Model 2 provides a better fit compared to Model 1 according to all the three fit measures. This suggests that, for the prediction of the purchase patterns

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|}
\hline
Model & Effect of shopping baskets on the arrival rates & Interdependence across category choices & Temporal dependence across category choices & Number of model parameters & log marginal density & In-sample MAE & Out-of-sample MAE \\
\hline
Model 1 & × & × & × & 16 & −94,161 & 0.66 & 0.30 \\
Model 2 & × & × & × & 27 & −93,946 & 0.62 & 0.28 \\
Model 3 & × & × & × & 40 & −93,853 & 0.61 & 0.27 \\
Model 4 & × & × & × & 47 & −93,742 & 0.60 & 0.26 \\
Model 5 & × & × & × & 51 & −93,178 & 0.57 & 0.24 \\
Model 6 & 1 dim. space & 1 dim. space & & 43 & −93,828 & 0.61 & 0.27 \\
Model 7 & Main effects only & & & 54 & −93,171 & 0.57 & 0.24 \\
Model 8 & & & & 72 & −92,413 & 0.51 & 0.21 \\
Model 9 & 1 dim. space & 2 dim. space & & 56 & −93,094 & 0.57 & 0.24 \\
Model 8 & 2 dim. space & 3 dim. space & & 59 & −93,802 & 0.54 & 0.22 \\
Model 8 & 3 dim. space & 4 dim. space & & 61 & −92,602 & 0.52 & 0.21 \\
Model 8 & 4 dim. space & & & 62 & −92,543 & 0.52 & 0.21 \\
\hline
\end{tabular}
\caption{Model comparison.}
\end{table}

\(^{10}\) Because Models 4.5 and 7 do not employ the latent space approach in specifying the arrival rate, the models also do not use the Euclidean norm in reparameterizing \(\theta_{ijk}\) and/or \(\psi_{ijk}\) in Eq. (10).

\(^{11}\) We were unable to compute the mean absolute percentage error (MAPE = |Predicted − Observed|/Observed) because the observed numbers of purchases of some shopping baskets were zero for some customers.
of shopping baskets, our approach of modeling the firm-level arrival process and category choice decisions conditional on a transaction occurring outperforms applying independent BG/NBD models to individual categories, even when we do not explicitly consider any of the associations in our proposed model. We find that Model 3 provides a better fit compared to Model 2, suggesting that the covariates considered in our study are useful in predicting customers’ purchase patterns across multiple categories. The better fit of Model 4 over Model 3 implies the importance of incorporating the temporal dependence across category choices. Comparing the performance of Model 5 to Model 4 suggests that customers’ choice decisions are correlated across product categories.

We find that Model 6 provides a better fit compared to Model 3, suggesting an association between the product categories comprising the shopping basket and the arrival rate. The better fit of Model 9 over Model 5 also implies the association of the two model components. In examining the performance of Models 6 and 9 with different dimensional latent spaces, we find that the model fit improves as we include additional dimensions in the space. However, we find negligible differences in model fit between the model with a three-dimensional space and the model with a four-dimensional space, implying that the flexibility gained from the additional fourth dimension is marginal. Comparing the performance of Model 7 to Model 5 suggests that considering the main effects of the categories on the arrival rate is helpful in predicting customers’ purchase patterns across categories. Furthermore, the better performance of Model 9 with higher-dimensional spaces, compared to Model 7, indicates that incorporating the effect of the overall shopping basket composition allows additional benefits beyond simply considering the main effects of the categories.

We find that the fit of Model 8 is slightly better than Model 9 in the calibration period but there is no difference between their performances in the validation period, suggesting that Model 9 successfully approximates the correlations within and across model components with a smaller number of parameters. Given that Model 9 has comparable forecasting ability to Model 8 with fewer parameters and that the latent space modeling approach more easily scales to accommodate more categories, we proceed to present detailed findings under this model specification.

In sum, our results suggest the importance of accounting for the association of shopping basket composition with the arrival rate, the cross-category effects within a category, and the temporal dependence across category choices. Compared to the benchmark model that omits all three sources of correlation (Model 3), the MAE for Model 9 is 15% lower in the calibration period and 22% lower in the validation period. As such, we focus the remainder of our discussion on the results from our proposed model with a three-dimensional latent space because the parsimony with a smaller number of model parameters outweighs the benefit of the minor improvement in the model fit with a four-dimensional space.

5.3. Model performance

To validate the model, we estimate the following measures, averaged across MCMC iterations: (1) the number of transactions, (2) the mean interarrival times after purchasing a specific shopping basket, and (3) the number of purchases of a shopping basket over the data period.

Using the estimates of the model parameters, we predict the customers’ transaction patterns at the firm level over time. Fig. 3 compares the predicted posterior means of the cumulative number of transactions to the corresponding observed values, aggregated across the customers. We find that the observed values are contained in the 95% posterior intervals of their respective predicted values throughout the data period. At the end of the calibration (validation) period, the model predicts the observed number at a 3.0% (2.6%) error rate, indicating that the model can accurately track the customers’ cumulative transactions.

We next compare the observed and predicted values of the mean interarrival times after the customers purchase each combination of categories in Fig. 4. The results show that all of the observed values are contained in the 95% posterior intervals of their respective predicted values except the case of the shopping basket {1,0,0,1}, which is among the least purchased shopping baskets. The average of the percentage errors weighted by the number of observations for shopping baskets is 8.9%. As expected, the prediction error tends to be smaller for more frequently observed shopping baskets and larger for less frequently observed ones (see Table 2 for the frequency of each shopping basket). The dotted lines in Fig. 4 represent the 95% posterior intervals of the predicted mean interarrival times under Model 5. Without considering the association of the interarrival process with shopping basket composition, the predicted interarrival times would not vary across shopping baskets. As a result, on average the benchmark model underpredicts the interarrival times after purchasing smaller shopping baskets and overpredicts the interarrival times after purchasing larger shopping baskets. Moreover, the observed values for more than half of the possible shopping basket compositions are not contained in the corresponding 95% posterior intervals. We find that the percentage error under Model 5 (Model 7) is 11.0% (10.8%), an increase in the percentage error of 24% (21%). This suggests that incorporating the effect of shopping basket composition on the arrival process improves firm-level predictions of transactional patterns.

Our proposed model also allows one to forecast purchases of customers’ shopping baskets. Marketers can use Eq. (8) to compute the probability that a customer purchases a specific shopping basket at a given transaction conditional on her past purchase records, which can aid firms in their promotional strategies. For example, to drive store traffic, marketers may incorporate those product categories in which a customer is most likely to purchase in targeted communications. Having identified those product categories that customers are most likely to purchase together may also aid retailers with in-store promotional activity.

To ensure that the model captures the customers’ purchase patterns of shopping baskets, we predict the cumulative number of purchases of each shopping basket over the data period. Fig. 5 compares the predicted posterior means of the cumulative number of transactions to the corresponding observed values, aggregated across the customers. We find that for all 15 different shopping baskets, the observed values are contained in the 95% posterior intervals of their respective predicted values throughout the data period. The tracking plots suggest that our model tracks the customers’ purchases of shopping baskets fairly well.

---

12 This may not be surprising because a good in-sample fit of more complicated models does not necessarily lead to a good holdout fit, often due to an overfitting problem (Malthouse, 1999).

13 We also tested and compared the forecasting ability of the proposed models with a three- and four-dimensional space, respectively, and found no significance difference in the performance between the models. Results are available from the authors on request.
At the end of the calibration (validation) period, the model predicts the cumulative purchases at a 7.2% (8.6%) error rate, averaged across shopping baskets. In comparison, the prediction errors increase to 10.2% (11.9%) under Model 4, corresponding to an increase in the error by 42% (38%). We find that Model 4 overestimates the purchases of shopping baskets containing only one category and underestimates the purchases of shopping baskets consisting of multiple categories as a result of ignoring the cross-category dependence detected under our proposed model. By incorporating the model component, Model 5 has lower errors compared to Model 4 (8.1% in calibration period and 9.3% in validation period).

Fig. 4. Model prediction on mean interarrival time after purchasing a shopping basket.

Fig. 5. Model prediction on cumulative purchases of a shopping basket.

Note: (1) In all charts, the x-axis is week and the y-axis is the cumulative purchases of a shopping basket.
(2) The solid and dotted lines represent the observed and predicted values, respectively.
in validation period). The results therefore highlight the importance of considering interdependence across category choices in evaluating customers’ likelihood of purchasing a specific combination of categories and thus forecasting cross-category sales over a future time period.

Lastly, by aggregating the customers’ purchase patterns of shopping baskets, we predict the cumulative category purchases over the data period. Fig. 6 shows the comparison of the distributions of the observed and predicted values across the customers for each category. The model fit results clearly support the ability of our model to capture the purchases of individual categories.

### 5.4. Parameter inferences

We describe inferences based on the posterior distributions of our model parameters. We begin with the results for the model of customer arrivals and attrition. From the estimates of $\mu_i$, $\mu_0$, $\Sigma_{\lambda\lambda}$, and $\Sigma_{\omega\omega}$ in Table 4a, we find that the mean baseline arrival rate is 0.10 ($= e^{\mu_0+2\lambda_0/2}$), and on average customers defect with a probability of 0.01 after their each transaction. Directionally, we find that the estimate of the coefficient for $LIT_i$ is negative, suggesting that a long lagged interarrival time reduces the purchase frequency, but this effect does not significantly differ from 0.

In Table 4b, we report the estimated coordinates of the four categories in the three-dimensional latent space. As specified in Eq. (4), the coordinates jointly determine the Euclidean norm for a chosen shopping basket which governs the subsequent interarrival times (see Fig. 4 for the predicted mean interarrival times after purchasing each shopping basket). Fig. 7 visualizes the position of the categories in the latent space, showing a set of two-dimensional maps to portray the three-dimensional latent space. From the first two panels of the figure, we see that basics and makeups are relatively close to each other with respect to their first coordinates, as are creams and serums. As such, the first dimension of the latent space may reflect characteristics of the categories which contrast basics and makeups to creams and serums in women’s use of cosmetics, such as “essential versus optional” or “basic versus functional.” From the third panel of the figure, we find that the second (third) dimension allows for differentiation between creams (serums) and the other product categories.

A next set of inferences arises from the multi-category choice model. Table 5a reports the parameter estimates that capture the category-specific effects specified in Eq. (7). Based on the estimates of $\mu_{ij}$’s, we find that customers’ average baseline purchase tendencies are highest for creams and lowest for makeups. The parameter estimates for $C_{ij-1}$’s indicate that the purchase of basics, creams and serums in the prior transaction reduces the likelihood of purchasing in those categories at the current transaction. In contrast to these categories, the prior purchase of makeups does not significantly affect the likelihood of purchasing the category. In all categories, the parameter estimates for both $ET_{ijk}$’s and $(ET_{ijk})^2$’s are significant, suggesting an inverse U-shaped relationship between the time since the last purchase of the category and the probability of purchasing the category. This indicates an increasing purchase probability up to some point, after which customers are increasingly less likely to purchase the category. Combined with the results of the timing model, this implies that a customer’s shopping timing and category choice decisions are sequentially associated with each other across transactions, an association that has not

### Table 4a

Parameter estimates of the timing and attrition models.

<table>
<thead>
<tr>
<th>Coordinates</th>
<th>Posterior mean</th>
<th>95% posterior interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basics</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Creams</td>
<td>0.38</td>
<td>[0.25, 0.50]</td>
</tr>
<tr>
<td>Serums</td>
<td>0.24</td>
<td>[0.17, 0.31]</td>
</tr>
<tr>
<td>Makeups</td>
<td>0.75</td>
<td>[0.60, 0.90]</td>
</tr>
</tbody>
</table>

### Table 4b

Parameter estimates of the timing model: coordinates of categories in a three-dimensional latent space.

<table>
<thead>
<tr>
<th>Coordinates</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basics</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Creams</td>
<td>0.38</td>
<td>0.97</td>
<td>0.00</td>
</tr>
<tr>
<td>Serums</td>
<td>0.34</td>
<td>0.35</td>
<td>0.91</td>
</tr>
<tr>
<td>Makeups</td>
<td>0.75</td>
<td>0.08</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Note: The numbers in brackets denote the 95% posterior interval.
previously been discussed in the customer relationship management literature.

In Table 5b, we report the estimates of parameters which jointly capture the interdependence across category choices, reflected by $\theta_{ikl}$. We find that the estimates of $\xi_{ikl}$ for all categories except makeups are significant. The estimate of $\xi_{1}$ is not significant, which implies that the cross-category effect in the multi-category choice process is not correlated with the relationship of categories in driving interarrival times. Using the parameter estimates, we compute the values of $\theta_{ij}$’s according to the specification in Eq. (9) across MCMC iterations, and summarize the results in Table 5c. With the exception of a pair of basics and makeups, the remaining cross-category effects are all positive and significant, which implies that purchasing in one category increases the probability of purchasing in the other category in the same transaction. Thus, the other five pairs of the categories are more likely to be purchased than a pair of basics and makeup. Among the five significantly associated pairs, creams and serums (creams and makeups) have the largest (smallest) positive impact on each other. Interestingly, we find that the cross-category effects involving makeups (the third column of Table 5c) tend to be smaller compared to the effects involving the other categories, suggesting a weaker association with the other categories. This may be due to makeups being relatively more distinct from the other three categories.

Table 5d presents the estimates of parameters which jointly capture the temporal dependence across category choices, reflected by $\psi_{ikl}$. From the estimates, the values of $\psi_{ikl}$’s are computed in Table 5e. The parameter estimates in the first column of the table reveal that the lagged purchase of basics has a significant and negative effect on the purchase of creams. From the estimates in the second (third) column, we find that the lagged purchase of creams (serums) has a significant and negative effect on the purchase of basics and serums (basics and creams). The estimates in the fourth column and row indicate that purchasing makeups on the prior trip adversely affects the probability to

<p>| Table 5a |</p>
<table>
<thead>
<tr>
<th>Parameter estimates of the choice model: category-specific effects.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Posterior mean</strong></td>
</tr>
<tr>
<td>Basics</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<p>| Table 5b |</p>
<table>
<thead>
<tr>
<th>Parameter estimates of the choice model: cross-category effects.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Posterior mean</strong></td>
</tr>
<tr>
<td>$\xi_{11}$</td>
</tr>
<tr>
<td>$\xi_{12}$</td>
</tr>
<tr>
<td>$\xi_{13}$</td>
</tr>
<tr>
<td>$\xi_{14}$</td>
</tr>
<tr>
<td>$\xi_{15}$</td>
</tr>
</tbody>
</table>

<p>| Table 5c |</p>
<table>
<thead>
<tr>
<th>Estimated $\Theta$.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Posterior mean</strong></td>
</tr>
<tr>
<td>Basics</td>
</tr>
<tr>
<td>Creams</td>
</tr>
<tr>
<td>Serums</td>
</tr>
<tr>
<td>Makeups</td>
</tr>
</tbody>
</table>

Note: The numbers in brackets denote the 95% posterior interval.

<p>| Table 5d |</p>
<table>
<thead>
<tr>
<th>Parameter estimates of the choice model: lagged cross-category effects.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Posterior mean</strong></td>
</tr>
<tr>
<td>$\xi_{11}$</td>
</tr>
<tr>
<td>$\xi_{12}$</td>
</tr>
<tr>
<td>$\xi_{13}$</td>
</tr>
<tr>
<td>$\xi_{14}$</td>
</tr>
<tr>
<td>$\xi_{15}$</td>
</tr>
<tr>
<td>$\xi_{16}$</td>
</tr>
<tr>
<td>$\xi_{17}$</td>
</tr>
<tr>
<td>$\xi_{18}$</td>
</tr>
<tr>
<td>$\xi_{19}$</td>
</tr>
</tbody>
</table>

<p>| Table 5e |</p>
<table>
<thead>
<tr>
<th>Estimated $\Psi$.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Posterior mean</strong></td>
</tr>
<tr>
<td>Basics</td>
</tr>
<tr>
<td>Creams</td>
</tr>
<tr>
<td>Serums</td>
</tr>
<tr>
<td>Makeups</td>
</tr>
</tbody>
</table>

Note: The numbers in brackets denote the 95% posterior interval.
purchase other categories while lagged purchases of other categories do not affect purchases of makeups. These results illustrate the modeling framework's flexibility in capturing an asymmetric relationship in the cross-category associations across transactions. Such insights can be of value to sales forces and marketers in tailoring how they approach customers, whether with cross-selling offers and customized communications, based on their past purchases.

Finally, we report the estimated variance–covariance matrix Σ in Table 6. The first row of the matrix shows the covariance between log λ and the intercepts for the other model components. The negative relationship between log λ and β indicates that customers who purchase more frequently are less likely to become inactive. We also find a negative relationship between log λ and β, which suggests that the shopping baskets of more (less) frequent shoppers are likely to be smaller (larger) in the number of categories. The result indicates the need to consider the association of the interarrival process and shopping basket composition across customers as well as across transactions within individuals.

6. Conclusion and future research

This research develops a model for multi-category customer base analysis in a noncontractual setting. We jointly model customers’ arrival process to the firm, multi-category choice decisions, and latent attrition in an integrated framework, allowing for shopping dynamics due to the interplay of purchase timing and incidence across categories. To capture the association between the arrival rate and shopping basket choice parsimoniously, we propose a novel modeling approach using a latent space of product categories. Our model also accounts for the independent choices of multiple categories and the correlation of repeatedly observed outcomes both within and across categories. Applying our model to category-level customer transaction data, we demonstrate the importance of accounting for the interplay of interarrival times and shopping basket composition, and the cross-category effects within and across transactions. The proposed model offers superior fit and performance in predicting customers’ purchase patterns across categories compared to several benchmark models. The ability of our model to forecast the purchase patterns of individual categories can assist marketers in managing relationships with customers. Our modeling approach goes beyond extant “buy ‘til you die” models that focus on a firm-level or a single type of transactional activity by enabling firms to assess customers’ likely future behavior at the more granular category level. Should firms be interested in making decisions based on customers’ behavior across multiple categories, the category-level results from our framework can be aggregated. Our model also allows marketers to identify which customers are more likely to purchase in a given set of categories on a particular visit, enabling the firm to dynamically tailor its marketing efforts to customers based on their purchase history in the categories.

There are several limitations in the current work that can be addressed in future research. First, given the limited extent of latent customer attrition inferred in our data, we model customer attrition at the firm level and employ the time-invariant attrition process. However, in other empirical contexts where customer attrition is more prevalent, customer attrition may vary across transactions depending on shopping basket composition. In particular, Schweidel, Park, and Jamal (2014) provide a multi-activity latent attrition model for customer base analysis. Their empirical findings suggest that the latent attrition processes are related, indicating that conducting one type of activity is informative of whether customers are still “alive” to conduct another type of activity, and consequently impacts inferences of customer value. It would be fruitful to incorporate different sources of attrition into our proposed model and untangle their effects on future behavior. Second, one could extend our modeling framework by modeling customers’ purchase amount decisions across categories. As our model extends the “buy ‘til you die” framework commonly used for customer valuation to a multivariate context, incorporating the amount model would allow one to quantify the contribution of individual categories to customer lifetime value (CLV) and assess the relationship between multi-category shopping behavior and CLV, which could assist the firm in allocating marketing resources across categories within individual customers.

Third, given the focus of this research and due to the limitation of the data, we have not considered the impact of marketing mix variables on customers’ shopping behavior. In practice, firms often implement several concurrent marketing programs at different levels. Some of them may be store-wide campaigns which affect both purchase timing and choice decisions across all categories (albeit with different degrees of effectiveness), while others focus on specific product categories and influence purchase behavior in not only promoted ones but also related others. To assess the impacts of such different marketing efforts on customers’ purchase behavior and thus the firm’s revenue, the interplay between the interarrival times and shopping basket composition as well as the cross-category dependence, needs to be considered. For example, store-wide campaigns may accelerate customers’ shopping frequency. As a result of the increased interarrival time, however, their probabilities of purchasing categories on a given trip may decrease. Similarly, several concurrent category-level promotions may help increase the probability of category choice. However, the resulting large shopping baskets could lead to a longer interarrival time until the customers’ next shopping trip. Failing to consider such interactions could result in erroneous estimates of the impact of the firm’s marketing activities. While we do not have access to such data in this empirical application, our modeling framework provides a general platform for the inclusion of different types of marketing covariates and thus allows one to disentangle their effects on customers’ purchase patterns across categories. Our model can be also modified to bring in the methods proposed by Gupta (1991), who demonstrated a sophisticated way of incorporating time-varying marketing mix variables and customers’ product inventory

### Table 6

<table>
<thead>
<tr>
<th></th>
<th>log λ</th>
<th>ω_0</th>
<th>β_{λ1}</th>
<th>β_{λ2}</th>
<th>β_{λ3}</th>
<th>β_{λ4}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.31</td>
<td>−0.05</td>
<td>−0.08</td>
<td>−0.07</td>
<td>−0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.23, 0.38]</td>
<td>[−0.09, −0.01]</td>
<td>[−0.11, −0.05]</td>
<td>[−0.10, −0.04]</td>
<td>[−0.01, 0.08]</td>
<td></td>
</tr>
<tr>
<td>ω_0</td>
<td>0.21</td>
<td>0.29</td>
<td>0.24</td>
<td>0.26</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.05, 0.38]</td>
<td>[0.07, 0.50]</td>
<td>[0.09, 0.50]</td>
<td>[0.16, 0.34]</td>
<td>[−0.02, 0.39]</td>
<td></td>
</tr>
<tr>
<td>β_{λ1}</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.12</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.17, 1.04]</td>
<td>[0.21, 0.39]</td>
<td>[0.24, 0.37]</td>
<td>[0.04, 0.21]</td>
<td>[0.18, 0.38]</td>
<td></td>
</tr>
<tr>
<td>β_{λ2}</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
<td>0.24</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.78, 1.04]</td>
<td>[0.78, 1.04]</td>
<td>[0.78, 1.04]</td>
<td>[0.15, 0.34]</td>
<td>[0.27, 0.47]</td>
<td></td>
</tr>
<tr>
<td>β_{λ3}</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.24</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.76, 1.02]</td>
<td>[0.76, 1.02]</td>
<td>[0.76, 1.02]</td>
<td>[0.15, 0.34]</td>
<td>[0.27, 0.47]</td>
<td></td>
</tr>
<tr>
<td>β_{λ4}</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.24</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.61, 1.118]</td>
<td>[0.61, 1.118]</td>
<td>[0.61, 1.118]</td>
<td>[0.15, 0.34]</td>
<td>[0.27, 0.47]</td>
<td></td>
</tr>
</tbody>
</table>

Note: The numbers in brackets denote the 95% posterior interval.
levels into a multi-event timing model. We hope that this study generates further interest and accelerates the progress in this important area of research.

Appendix A. Derivation of Eq. (8)

Let us denote the probability of the “no-purchase” case derived using Eq. (6) as

\[
\Pr\left( C_{ij} = 0 \right) = \Pr\left( C_{ij} = 0, C_{ij} = 0, \ldots, C_{ij} = 0 \right).
\]  
(A.1)

Then, from the theorem by Besag (1974), we have

\[
\begin{align*}
\Pr\left( C_{ij} = C_{i2} = \ldots = C_{ij} = 0 \right) &= \prod_{k=0}^{0} \Pr\left( C_{ik} = C_{ik} = C_{ik} = \ldots = C_{ij} = 0 \right) \\
&= \prod_{k=0}^{0} \Pr\left( C_{ik} = C_{ik} = C_{ik} = \ldots = C_{ij} = 0 \right) \\
&= \prod_{k=0}^{0} \Pr\left( C_{ik} = C_{ik} = C_{ik} = \ldots = C_{ij} = 0 \right) \\
&= \prod_{k=0}^{0} \Pr\left( C_{ik} = C_{ik} = C_{ik} = \ldots = C_{ij} = 0 \right) \\
&= \prod_{k=0}^{0} \Pr\left( C_{ik} = C_{ik} = C_{ik} = \ldots = C_{ij} = 0 \right).
\end{align*}
\]  
(A.2)

By substituting Eq. (6) into Eq. (A.2), we have

\[
\begin{align*}
\Pr\left( C_{ij}, C_{i2}, \ldots, C_{ij} \right) &= \Pr\left( C_{ij} = C_{ij} = C_{ij} = \ldots = C_{ij} = 0 \right) \\
&= \Pr\left( \exp\left( \pi_{ij} + \Theta_{i}C_{ij} + \Psi_{j}(1-\pi_{ij}) \right) \right) \\
&= \Pr\left( \exp\left( \pi_{ij} + \Theta_{i}C_{ij} + \Psi_{j}(1-\pi_{ij}) \right) \right) \\
&= \Pr\left( \exp\left( \pi_{ij} + \Theta_{i}C_{ij} + \Psi_{j}(1-\pi_{ij}) \right) \right) \\
&= \Pr\left( \exp\left( \pi_{ij} + \Theta_{i}C_{ij} + \Psi_{j}(1-\pi_{ij}) \right) \right). \\
\end{align*}
\]  
(A.3)

where \( \Theta_{i} \) and \( \Psi_{j} \) are the \( k \)th row vector of matrices \( \Theta \) and \( \Psi \), respectively, and

\[
C_{ij}^{0,23, \ldots} = \{0, 0, 0, \ldots, C_{ij} \} \quad \text{and} \quad C_{ij}^{0,0,0, \ldots} = \{0, 0, 0, \ldots, C_{ij} \}.
\]

Eq. (A.3) can be written as:

\[
\begin{align*}
\Pr\left( C_{ij}, C_{i2}, \ldots, C_{ij} \right) &= \Pr\left( C_{ij} = C_{ij} = C_{ij} = \ldots = C_{ij} = 0 \right) \\
&= \Pr\left( \exp\left( \pi_{ij} + \Theta_{i}C_{ij} + \Psi_{j}(1-\pi_{ij}) \right) \right) \\
&= \Pr\left( \exp\left( \pi_{ij} + \Theta_{i}C_{ij} + \Psi_{j}(1-\pi_{ij}) \right) \right) \\
&= \Pr\left( \exp\left( \pi_{ij} + \Theta_{i}C_{ij} + \Psi_{j}(1-\pi_{ij}) \right) \right) \\
&= \Pr\left( \exp\left( \pi_{ij} + \Theta_{i}C_{ij} + \Psi_{j}(1-\pi_{ij}) \right) \right). \\
\end{align*}
\]  
(A.4)

By arranging exponential terms in Eq. (A.4), we have

\[
\begin{align*}
\Pr\left( C_{ij}, C_{i2}, \ldots, C_{ij} \right) &= \Pr\left( C_{ij} = C_{ij} = C_{ij} = \ldots = C_{ij} = 0 \right) \\
&= \Pr\left( \exp\left( \pi_{ij} + \Theta_{i}C_{ij} + \Psi_{j}(1-\pi_{ij}) \right) \right) \\
&= \Pr\left( \exp\left( \pi_{ij} + \Theta_{i}C_{ij} + \Psi_{j}(1-\pi_{ij}) \right) \right) \\
&= \Pr\left( \exp\left( \pi_{ij} + \Theta_{i}C_{ij} + \Psi_{j}(1-\pi_{ij}) \right) \right). \\
\end{align*}
\]  
(A.5)

Because the sum of the joint probability in Eq. (A.5) across all possible combinations of categories should be equal to one and our data do not include the “no-purchase” case, the joint distribution of \( C_{ij} = \{C_{i1}, C_{i2}, \ldots, C_{ij} \} \) is given by:

\[
\begin{align*}
P\left( C_{ij} = \{C_{ij}, \ldots, C_{ij} \} | 1 \right) &= \exp\left( \pi_{ij} C_{ij} + \frac{1}{2} C_{ij} \Theta C_{ij} + \pi_{ij-1} C_{ij-1} \right) \\
&= \sum_{C_{ij} = 0}^{\infty} \exp\left( \pi_{ij} C_{ij} + \frac{1}{2} C_{ij} \Theta C_{ij} + \pi_{ij-1} C_{ij-1} \right).
\end{align*}
\]  
(A.6)
Full Length Article

The performance implications of outsourcing customer support to service providers in emerging versus established economies

Néomie Raassens a,⁎, Stefan Wuyts b,c,1, Inge Geyskens c,2

a Eindhoven University of Technology, The Netherlands
b Koç University, Turkey
c Tilburg University, The Netherlands

ABSTRACT

Recent discussions in the business press query the contribution of customer-support outsourcing to firm performance. Despite the controversy surrounding its performance implications, customer-support outsourcing is still on the rise, especially to emerging markets. Against this backdrop, we study under which conditions customer-support outsourcing to providers from emerging versus established economies is more versus less successful. Our performance measure is the stock-market reaction around the outsourcing announcement date. While the stock market reacts, on average, more favorably when customer-support is outsourced to providers located in emerging markets as opposed to established economies, approximately 50% of the outsourcing firms in our sample experience negative abnormal returns. We find that the shareholder-value implications of customer-support outsourcing to emerging versus established economies are contingent on the nature of the customer support that is being outsourced and on the nature of the outsourcing firm. Customer-support outsourcing to emerging markets is less beneficial for services that are characterized by personal customer contact and high knowledge embeddedness than for customer-support services that involve impersonal customer contact and are low on knowledge embeddedness. Firms higher in marketing resource intensity and larger firms benefit more from outsourcing customer-support services to emerging markets than firms lower in marketing resource intensity and smaller firms.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Increasingly, firms are outsourcing customer support. A research report by IDC (2011) reveals that the worldwide outsourced customer-care services market will grow at a compound annual growth rate of 5% to reach $66.2 billion in 2015. While the large majority of customer-support outsourcing is still done locally (Mudambi & Venzin, 2010), firms in established economies are increasingly considering emerging markets as attractive outsourcing locations (Javalgi, Dixit, & Scherer, 2009). Recent examples include Barclays, a global financial service provider that outsources call-center jobs to India, and T-Mobile UK that outsources part of its customer care to the Philippines.

Although customer-support outsourcing is all the rage, many outsourcing arrangements are not successful. Anecdotal accounts in the business press suggest that organizations can achieve cost savings of 25 to 30% through customer-support outsourcing (Gartner, 2005), especially when outsourcing to emerging markets with lower labor costs (Bharadwaj & Roggeveen, 2008). However, blinded by quick-fix cost savings, many firms forget that there can also be “hidden costs” in outsourcing services (Ren & Zhou, 2008, p. 370). Deloitte Consulting (2005, p. 21) states:

“50 percent of outsourcing in the near future will be successful, with the failures stemming from clients that don’t know what they are doing, don’t understand outsourcing, or don’t understand their own business.”

Against this backdrop, the goal of this research is to understand what distinguishes more versus less successful customer-support outsourcing practices.

We contribute to the extant literature in three ways. First, we study the financial performance implications of outsourcing customer support. Recent studies by Thelen and colleagues have focused on the negative consequences of customer-support outsourcing for consumers’ service quality perceptions (Thelen, Honeycutt, & Murphy, 2010), sentiments (Thelen, Yoo, & Magnini, 2011), and subsequent (adverse) behaviors toward the outsourcing firm (Thelen & Shapiro, 2012). In contrast, when executives are asked about the financial impact of their outsourcing
outsourcing and outline our conceptual framework. Then, we introduce the literature on the performance implications of customer-support outsourcing to service providers from emerging versus established economies. Although a growing body of literature has examined customer-support outsourcing, most of this research has not distinguished between service providers from established versus emerging markets (e.g., Ren & Zhou, 2008; Thelen et al., 2011). The recent spread in offshore outsourcing to emerging markets makes it paramount for marketing scholars to shift their attention. As such, we address the call for more research on marketing in emerging markets (Burgess & Steenkamp, 2013; Dekimpe, 2009).

Third, there has been limited research to date that examines how the performance implications of outsourcing customer support to emerging versus established economies may differ across firms. We examine whether these performance implications depend on (1) the nature of the customer-support service and the nature of the outsourcing firm that is being outsourced, and (2) the nature of the firm that is outsourcing the customer-support activities. Specifically, emerging markets present significant departures from established economies in terms of culture as well as regulation (Burgess & Steenkamp, 2006), which may lead to communication-style problems and enforcement problems, respectively. We argue that the extent to which outsourced services are prone to communication-style problems depends on the type of customer support that is being outsourced. Further, the extent to which firms are vulnerable to regulatory problems depends on the nature of the firm that is outsourcing. By comparing the financial performance implications of customer-support outsourcing to providers from emerging versus established economies, and by identifying the factors that distinguish more versus less successful outsourcing practices, we hope to assist executives in avoiding future costly mistakes.

The remainder of this study is organized as follows. We first review the literature on the performance implications of customer-support outsourcing and outline our conceptual framework. Then, we introduce the hypotheses, describe the methodology, and present the results. In the final section, we discuss the study’s theoretical and managerial implications and we outline avenues for future research.

2. Theory

2.1. Literature review

Marketing strategy research has long emphasized the importance of studying the performance consequences of customer-support outsourcing. This has fueled a significant, multifaceted literature (see Table 1 for an overview). A first stream of literature has analyzed the performance implications of customer-support outsourcing for the outsourcing firm analytically. Hasija, Pinker, and Shumsky (2008) and Ren and Zhou (2008) examine how different contracts enable the outsourcing firm to increase its profits. Milner and Olsen (2008) study how alternate contracts affect service quality in the form of minimal customer delays. None of these studies consider how the performance consequences of customer-support outsourcing may differ across providers from emerging versus established economies.

In addition to the game-theoretical literature, a number of studies have tested the performance implications of customer-support outsourcing empirically. These studies can be described along two dimensions: (1) the nature of the performance measure used, and (2) the geographic scope of the study.

Most studies have examined the impact of customer-support outsourcing on consumer perceptions, sentiments, and behavioral intentions. Thelen et al. (2010) show that U.S. citizens’ perceptions of service quality vary considerably by the outsourcing provider’s country, with providers from established economies (e.g., Canada, Ireland) scoring much higher on perceived service quality than providers from emerging markets (e.g., China, India, Philippines). Roggeveen, Bharadwaj, and Hoyer’s (2007) experiments among U.S. MBA students indicate that locating a call center in India or the Philippines as opposed to England impacts anticipated satisfaction negatively, but only for outsourcing firms that are unknown (as opposed to reputable). In a similar vein, Bharadwaj and Roggeveen (2008) find that U.S. customers experience greater satisfaction with a call center representative’s communication skills and problem solving ability when a call center is located domestically rather than in India.

Other studies corroborate the finding that consumers harbor negative feelings toward services that are outsourced offshore (i.e., to a country other than the outsourcing firm’s domestic market). They find that U.S. consumers’ negative sentiments towards offshore outsourcing decrease their satisfaction with the service received (Sharma, 2012; Sharma, Mathur, & Dhawan, 2009) as well as their attitude and commitment towards the offshoring firm (Thelen et al., 2011). Consumers’ negative sentiments towards offshoring further affect their behavioral intentions, ranging from decreased positive word-of-mouth intentions and increased intentions to boycott the outsourcing firm (Thelen et al., 2011) to requests for domestic service providers and intentions to stop doing business with the firm (Thelen & Shapiro, 2012). These studies, however, do not compare providers from emerging economies with providers from established economies. For example, Whittaker, Krishnan, and Fornell (2008) contrast domestic outsourcing to offshoring. They show that customer-support outsourcing decreases customer satisfaction in both cases, but they do not consider potential differences between providers from emerging markets and (nondomestic) providers from established economies.

We were able to locate only one study that addresses the financial performance consequences of customer-support outsourcing. Kalaignanam, Kushwaha, Steenkamp, and Tuli (2013) study the effect on shareholder value of outsourcing two types of Customer Relationship Management (CRM) processes—presales CRM (i.e., lead generation and order fulfillment) and postsales CRM (i.e., customer support). Although Kalaignanam et al. (2013) do not explicitly compare the performance consequences of outsourcing to providers from emerging versus established markets, they do study the role of cultural distance between the outsourcing firm and the outsourcing provider. They find that outsourcing customer support (i.e., postsales CRM) to culturally distant outsourcing providers has no adverse effect on shareholder value.3

In sum, while firms increasingly outsource to service providers located in emerging markets, research that explicitly compares outsourcing to providers from emerging versus established markets remains scarce (Mudambi & Venzin, 2010). In addition, even though firms primarily pursue cost savings when outsourcing customer support to emerging markets (a supply-side advantage), the scant empirical academic research has almost exclusively focused on the negative consumer appraisals that may result from customer-support outsourcing (a demand-side disadvantage). While it is of interest to quantify the impact on various performance indicators separately, it is first and foremost important to understand the overall net performance impact of outsourcing to emerging versus established economies. Apart from quantifying this net effect, the current study develops a contingency framework for identifying the boundary conditions of the relationship between customer-support outsourcing to emerging versus established economies and the outsourcing firm’s performance.

3 Kalaignanam et al.’s (2013) study differs from our study in two important ways. First, whereas cultural distance is indeed an important dimension along which emerging markets and established economies differ, it is not the only one. As we will argue below, emerging markets are also markedly different from established economies in terms of institutional systems, which may create regulatory problems (Burgess & Steenkamp, 2006). Second, whereas Kalaignanam et al. (2013) analyze the performance consequences of outsourcing postsales CRM to culturally distant countries, we argue for a contingency perspective and hypothesize that the performance consequences of outsourcing customer support to emerging markets (rather than established economies) are contingent on the nature of the outsourced customer-support service and the nature of the outsourcing firm.
2.2. Conceptual framework

Albeit prior studies have used slightly different approaches to distinguish emerging markets from established economies, they have in common that the country's welfare per capita is the key distinguishing characteristic. We follow Burgess and Steenkamp (2006, p. 339) in adopting the World Bank's definition of emerging markets as "countries in which PPP-adjusted GDP per capita, converted to U.S. dollar and smoothed for three-year currency fluctuations, is equal to or less than the highest ranked country classified as 'middle income' by the World Bank."

Customer-support outsourcing to emerging markets can have both positive and negative effects on the performance of the outsourcing firm. On the positive side, firms may benefit from cost savings by taking advantage of lower labor costs offered by the outsourcing provider. On the negative side, the markedly different cultural and institutional systems of emerging markets (Burgess & Steenkamp, 2006) may create communication-style problems and regulatory problems, respectively. Communication-style problems may occur not only between the outsourcing provider and the outsourcing firm's customers (e.g., Stringfellow, Teagarden, & Nie, 2008; Thelen et al., 2010) but also between the outsourcing provider and the outsourcing firm's employees. These communication difficulties make it more challenging for the outsourcing firm to coordinate activities with outsourcing providers from emerging markets (cf. Simonin, 1999), thereby increasing the outsourcing firm's costs and/or concerns about its customer service standards and quality.

In addition, regulatory problems may occur because emerging markets are characterized by less effective regulatory systems and contract-enforcement mechanisms, while established economies are marked by more formal and transparent rules and restrictions (Khanna, Palepu, & Sinha, 2005). When firms outsource to countries with a strong regulatory system, they can expect their outsourcing providers to act upon the normative influence of regulation (Edelman & Suchman, 1997) to avoid penalties for noncompliance (Hoffman, 1999). This is a dangerous assumption to make when outsourcing to emerging markets. Due to emerging economies' poor regulative system, outsourcing firms are less confident that their outsourcing providers will adhere to local laws and contractual agreements (Khanna et al., 2005). This may further increase the outsourcing firm's costs and/or may negatively affect service quality.

The communication-style and regulatory challenges of customer-support outsourcing are characteristic for outsourcing to emerging markets but less so for outsourcing to established economies. Hence, we expect that the performance implications of outsourcing customer support to emerging versus established economies depend on (1) the extent to which the outsourced service is prone to communication difficulties, and (2) the extent to which the outsourcing firm is vulnerable to regulatory threats. If the outsourced services are less prone to communication-style problems and the firm's vulnerability to regulatory problems is lower, the performance implications of outsourcing customer support to emerging economies, compared to outsourcing to established economies, will be more favorable. We argue that the extent to which outsourced services are prone to communication-style problems is a function of the nature of the outsourced customer-support service, while the extent to which firms are vulnerable to regulatory challenges depends on the nature of the outsourcing firm. Fig. 1 summarizes our conceptual framework.

2.3. The nature of the outsourced customer-support service

We argue that the proneness of customer-support services to communication-style problems varies along two key dimensions (Youngdahl & Ramaswamy, 2008): (1) the nature of the contact between the service employee and the outsourcing firm's customers (impersonal versus personal), and (2) the level of knowledge embeddedness of the outsourced customer-support activity (simple services versus complex solutions). Whereas the former dimension captures potential communication-style problems between the outsourcing provider and the outsourcing firm's customers, the latter dimension reflects possible communication-style problems between the outsourcing provider and the outsourcing firm's employees.

2.3.1. Personal customer contact

Customer contact pertains to employee–customer interaction during the delivery of customer support (Hartline & Ferrell, 1996), and can be personal or impersonal. Customer contact is personal when the production and consumption of the service occur simultaneously through a channel that accommodates direct, real-time interaction between the service employee and the customer. For example, call centers involve personal contact between service employees and customers. In contrast, e-mail support services do not involve direct, real-time interaction and are therefore impersonal.

Communication is contextually bound (Newburry & Yakova, 2006). The low cultural proximity between firms from established economies and firms from emerging markets may create a barrier to interaction, a "friction-related invisible cost of communicating" (Stringfellow et al., 2008, p. 170). For example, directness in speech is prized in the U.S. but may be perceived as rudeness in China. In a similar vein, "yes" in India actually means "I've heard you" whereas U.S. customers associate "yes" with making a commitment, which can cause problems of overpromising (Metters, 2008). Hence, when firms from established economies outsource their customer support to providers from culturally dissimilar emerging markets, effective customer support delivery is more at risk than when they outsource to providers from culturally more proximate established economies.

Whereas communication skills are important in delivering any type of customer support, they are vital when the service requires personal customer contact (Thelen et al., 2011). We argue that outsourcing to emerging markets is relatively more prone to communication-style problems when customer contact is personal rather than impersonal in nature. The differences between the customers' and the provider's...
communication styles may be more difficult to get around when the service involves personal (direct, real-time) interaction than when it involves impersonal (indirect) interaction, because personal interaction leaves less room for inspection or revision (Patterson, Cowley, & Prasongsukarn, 2006). In contrast, asynchronous, non-face-to-face communication that relies on written messages is easier to code in a communication style that fits customers from established economies. We therefore hypothesize:

**H1.** The effect on shareholder value of outsourcing customer support to emerging markets (rather than established economies) is lower for customer-support activities that are personal in nature than for customer-support activities that are impersonal in nature.

### 2.3.2. Knowledge embeddedness

Outsourced customer-support services also have different levels of knowledge embeddedness. Knowledge embeddedness is defined as the extent to which service providers need specialized skills and expert knowledge to deliver the customer support according to specifications (Youngdahl & Ramaswamy, 2008). Low knowledge embeddedness characterizes simple services (e.g., reservations), whereas high knowledge embeddedness characterizes complex solutions (e.g., technical customer support). The skills and knowledge required for customer-support services characterized by low knowledge embeddedness are relatively easy for the outsourcing firm to transfer to its provider without substantial misunderstandings (Liu, Feils, & Scholnick, 2011; Madhavan & Grover, 1998). On the other hand, the specialized skills and expert knowledge associated with complex solutions are notoriously difficult to transfer and require close interaction (Madhavan & Grover, 1998). Aron and Singh (2005) point out that when outsourcing providers require a great deal of domain experience and deep product knowledge to execute task processes, “they are unlikely to get those processes right for a long time” (p. 138).

We argue that transferring embedded knowledge is even more difficult when outsourcing customer support to emerging markets as compared to established economies. The communication-style problems that we discussed earlier in the context of employee–customer interactions may also arise between the outsourcing firm’s employees and the outsourcing provider’s employees (Stringfellow et al., 2008, p. 173), rendering any knowledge transfer attempt to a provider from an emerging market less effective than to a provider from an established economy. These knowledge-transfer problems are further exacerbated when the provider’s trainees in turn train local employees underneath them. This is common practice when outsourcing to emerging markets, as Web.com, a medium-sized U.S.-based web hosting and Internet service provider, discovered when outsourcing customer service to Bangalore, India:

> “Another negative side effect that we experienced was what we’ve now coined a ‘replicate fade,’ meaning our service and best practices became a copy of a copy of a copy. The highly qualified and skilled people we trained for three months at our corporate headquarters were training people underneath them who were then training others, and so on. Because the effectiveness of our training program was being diluted, our customers were not receiving the stellar customer support that is the cornerstone of our business”.

We therefore hypothesize:

**H2.** The effect on shareholder value of outsourcing customer support to emerging markets (rather than established economies) is lower for customer-support activities with high knowledge embeddedness than for customer-support activities with low knowledge embeddedness.

### 2.4. The nature of the firm

The performance implications of outsourcing customer support to emerging versus established economies not only depend on the nature of the outsourced customer-support service, but also on the nature of the outsourcing firm. Some firms are better able than others to cope with regulatory problems that are prevalent in emerging markets. Below, we argue that an outsourcing firm’s vulnerability to regulatory problems in emerging markets is contingent on the outsourcing firm’s marketing resource intensity and on its size.

#### 2.4.1. Marketing resource intensity

Regulative institutions have an essential role in supporting the effective functioning of the market mechanism, such that firms and individuals can engage in transactions without incurring undue costs or risks (Meyer, Estrin, Bhaumik, & Peng, 2009). Whereas regulative institutions are strong in established economies, they tend to be weak in emerging markets (McMillan, 2008). Thus, when firms outsource their customer support to emerging markets, they are faced with less effective contract-enforcement mechanisms (Liu et al., 2011). Therefore, they can be less confident that their outsourcing providers will adhere to contractual agreements (Khanna et al., 2005).
Wathne and Heide (2000) have argued that superior partner selection may help firms address contract-enforcement problems. However, whereas strong institutions come with a transparent disclosure environment, thereby providing firms with easy access to information about potential business partners and reducing information asymmetries, the weak institutional arrangements found in emerging markets typically amplify information asymmetries (Meyer et al., 2009). As a result, outsourcing firms need to spend more resources searching for partner-related information in emerging markets (Tong, Reuer, & Peng, 2008). Since there is no easy way of evaluating the quality of the partner beforehand, outsourcing firms may face an adverse selection problem.

Marketing resource intensity is the extent to which a firm invests in marketing activities (Mizik & Jacobson, 2007), such as advertising, branding, and customer service (Dinner, Mizik, & Lehmann, 2009). In the spirit of Day (1994), we argue that firms higher on marketing resource intensity have a stronger external focus, which enables them to better scan the environment. Their scanning ability allows them to better validate potential outsourcing providers’ claims and select a good partner. As a consequence, their cost of coping with regulatory challenges is reduced. We hypothesize:

**H3.** Firms with higher marketing resource intensity create more shareholder value by outsourcing customer support to emerging markets (rather than established economies) than firms with lower marketing resource intensity.

### 2.4.2. Firm size

The size of a firm is indicative of its financial resources (Contractor, Kumar, & Kundu, 2007; Cui & Lui, 2005), with larger firms having more financial resources than smaller firms (Johnson & Tellis, 2008). Hitt, Dacin, Levitas, Arregle, and Borza (2000) find that emerging-market firms attribute more importance to their alliance partners’ financial resources than firms in established economies. Given their attractiveness as exchange partners in emerging markets, larger outsourcing firms can therefore more credibly threaten to switch providers in case their current provider attempts to exploit the weak regulatory environment. As such, we expect that outsourcing providers will more likely adhere to contractual agreements if they are dealing with larger outsourcing firms. This will positively impact service quality, thereby increasing the outsourcing firm’s performance.

Moreover, larger outsourcing firms have more contact and clout in emerging markets than smaller firms (Acs, Morck, Shaver, & Yeung, 1997) and often receive preferential treatment from the local government (Cui & Lui, 2005; Schiller & Weder, 2001), because they offer social benefits such as employment creation (Child & Tsai, 2005). Their stronger ties with local governments in emerging markets provide outsourcing firms with better insight into local market conditions (Radjou & Prabhu, 2012), which may increase their ability to select a good exchange partner. In sum, larger firms can credibly threaten to switch providers and may be able to select better partners through their stronger ties with local governments, both of which increase their ability to deal with the regulative challenges inherent to outsourcing to emerging markets. We therefore hypothesize:

**H4.** Larger firms create more shareholder value by outsourcing customer support to emerging markets (rather than established economies) than smaller firms.

### 3. Methodology

#### 3.1. Evaluating the performance implications of customer-support outsourcing

Evaluating the performance implications of customer-support outsourcing is not straightforward. Commonly used performance measures such as return on assets, sales, or customer satisfaction may be less appropriate indicators to capture the performance implications of customer-support outsourcing, for the following reasons. First, performance is a multifaceted construct and examining any single performance facet in isolation—such as sales or customer satisfaction—is not likely to produce an adequate overall assessment (Kumar, Stern, & Achrol, 1992). This is particularly important in an outsourcing context, since the performance effects of outsourcing decisions can be wide-ranging and call for consideration of both demand-side and supply-side effects (Kalaigiannam & Varadarajan, 2012). On the demand side, customer-support outsourcing may affect, amongst others, customer satisfaction (Bharadwaj & Roggeveen, 2008), customer referrals (Thelen et al., 2011), and thereby customer sales. Supply-side effects refer to the impact of customer-support outsourcing on a firm’s cost structure (Kalaigiannam & Varadarajan, 2012). Customer-support outsourcing may not only lead to cost advantages through economies of scale and lower labor costs (Stringfellow et al., 2008), but also to more “hidden costs,” such as the costs of qualifying a good outsourcing provider, the costs of training the provider, and the costs of managing the provider to ensure that he lives up to his agreements (Business Week, 2003).

Second, traditional financial accounting measures such as return on assets have a historical orientation as opposed to a forward-looking focus. Yet, the financial outcomes of customer-support outsourcing decisions can be substantially delayed. Indeed, “companies that value short-term profit at the expense of meaningful customer service risk sacrificing long-term profits and the company’s own reputation” (Business Week, 2007). Because accounting measures only evaluate historical performance indicators, they are not well suited to capture such anticipated future cash flows (Srinivasan & Hansens, 2009).

Third, the temporal aggregation level of financial accounting measures makes the link to specific events or decisions difficult (Geyskens, Gielens, & Dekimpe, 2002). End-of-the-year accounting numbers may be influenced by many marketing and strategic decisions that took place during the year, of which the customer-support outsourcing decision is just one.

To deal with these three issues, we use an event study to examine the effect of outsourcing customer support on shareholder value. First, shareholder value integrates multiple facets of performance into a single “net-effect” measure (Gielens & Geyskens, 2012). Second, shareholder value hinges largely on growth prospects and sustainability of profits (i.e. how the firm is expected to perform in the future), in contrast to most accounting measures that are retrospective in examining historical performance (Srinivasan & Hansens, 2009). Third, an event study allows for measuring the impact of a specific, discrete event on daily (i.e. temporally disaggregated) stock returns, and thus can be thought of as a controlled experiment (Mizik & Jacobson, 2003).4

#### 3.2. Event-study methodology

The event-study approach relies on the assumption that financial markets are efficient. According to the semi-strong version of the efficient-market hypothesis, a firm’s stock price accurately reflects all publicly available information about the firm. When an event occurs (in our case, when information concerning a firm’s outsourcing of customer support is made public), investors update their expectations about the firm’s future performance and react by buying or selling shares of that firm. As a result, the firm’s stock price immediately changes to reflect the expected changes in future revenue streams associated with new information (Gielens, van de Gucht, Steenkamp, & Dekimpe, 2008). The percentage change in the stock price is the stock return.

We compare the observed stock return \(R_t\) on the event day (i.e. the day firm i’s outsourcing arrangement was announced) with \(E(R_t)\), the

---

4 Despite the advantages of using shareholder value as a performance metric, it is not perfect. Stock prices, just like accounting measures, can be manipulated.
firm's return that would be expected if the event had not taken place. The difference between the observed return for firm i on the event day and its expected return is the abnormal return, ARt, or the firm's unexpected change in stock price, which is attributed to the event. The abnormal return ARt provides an unbiased estimate of the future earnings generated by the event.

To obtain estimates of a firm's expected returns, we use the world-market model developed by Park (2004).\(^5\) According to this model:

\[
E(R_{it}) = \alpha_i + \beta_1 R_{mit} + \gamma_i R_{wmt} + \delta_i X_{it}
\]

(1)

where \(R_{mit}\) is the market-index return in the home country of outsourcing firm i on trading day \(t\). \(R_{wmt}\) is the world-market-index return on trading day \(t\), and \(X_{it}\) is the change in the foreign currency exchange rates in the home country of firm i on day \(t\). \(\alpha_i\), \(\beta_1\), \(\gamma_i\), and \(\delta_i\) are firm-specific OLS estimates from regressing \(R_t\) on \(R_{mit}\), \(R_{wmt}\), and \(X_t\) over an estimation period from 250 to 30 trading days prior to the event.

To account for information leakage before the event day (for \(t_i\) time periods before the event) and for the possibility that some information is disseminated after the event day (for \(t_x\) time periods after the event) (McWilliams & Siegel, 1997), we aggregate the abnormal returns for a firm over the event window \([-t_i, t_x]\) into a cumulative abnormal return \(CAAR\) to draw overall inferences for the event of interest (where \(t = 0\) on the event day):

\[
CAAR[-t_i, t_x] = \sum_{t=-t_i}^{t_x} AR_t
\]

(2)

Because we conduct the event study over \(N\) outsourcing events, this \(CAAR\) can be averaged into a cumulative average abnormal return \((CAAR)\):

\[
CAAR[-t_i, t_x] = \frac{1}{N} \sum_{i=1}^{N} CAAR[-t_i, t_x] / N
\]

(3)

To test the significance of the \(CAAR\), we use the Patell (1976) statistic, in which the abnormal returns are standardized by the standard deviations of the regression residuals that were obtained for the estimation window (cf. Jain, 1982). This reduces problems of heteroskedasticity that may arise when the estimated variances of the world-market index, trading day, and \(X_{it}\) change rates in the home country of outsourcing firm i on day \(t\).

To test our hypotheses on the performance consequences of outsourcing customer support, we regress the outsourcing firms' standardized cumulative abnormal returns on the covariates:

\[
CAAR[-t_i, t_x] = \beta_0 + \beta_1 Emerge + \beta_2 PersCont + \beta_3 KnowEmb + \beta_4 (PersCont \times Emerge) + \beta_5 (KnowEmb \times Emerge) + \beta_6 Mkthnt + \beta_7 Size + \beta_8 (Mktnt \times Emerge) + \beta_9 (Size \times Emerge) + \sum_{k=1}^{N} \gamma_k Control_{ik} + \mu_i
\]

(4)

where \(Emerge\) refers to the selection of an emerging market rather than an established economy as outsourcing location, \(PersCont\), and \(KnowEmb\), capture whether the outsourced customer support requires personal customer contact and embedded knowledge, respectively. \(Mktnt\) is the outsourcing firm's marketing resource intensity, \(Size\) indicates the outsourcing firm's size, and \(Control_{ik}\) is a set of \(K\) control variables. To account for potential inter-correlation among outsourcing firms located in the same continent, we use robust clustered error-term estimation (cf. Mizik & Jacobson, 2009).\(^5\)

Firms may have chosen an emerging market versus an established economy as an outsourcing location to optimize their performance, such that only maximizing choices are observed. To account for self-selection, we follow the Heckman two-step procedure (Heckman, 1979). In the first step, where we also take into account the clustered nature of the data, we specify a probit selection model to estimate the likelihood that a firm would engage in customer-support outsourcing to an emerging versus an established market (Emerge). As determinants, we include variables at the transaction, the firm, the industry, and the (home) country level which previous studies have identified as possible drivers of the likelihood of outsourcing to emerging markets. At the transaction level, we account for the two customer-support characteristics (personal customer contact and knowledge embeddedness) that are included in Eq. (4). Firms may be less inclined to outsource customer support that is personal in nature and that is characterized by high levels of embedded knowledge to a provider from an emerging market, for reasons outlined in the hypotheses section. At the firm level, we control for the two firm characteristics (marketing resource intensity and firm size) that are included in Eq. (4), again for reasons outlined in the hypotheses section. We further include the firm's past performance, as reflected in its profitability and leverage in the year prior to the outsourcing announcement. On the one hand, good past performance may motivate companies to experiment with outsourcing to providers from emerging markets. On the other hand, poor past performance signifies the ineffectiveness of existing operations and thus may provide strong and legitimate reasons for firms to outsource to providers from emerging markets in an attempt to obtain cost savings (Zhou, Tse, & Li, 2006). At the industry level, we include industry concentration. Highly competitive industries typically force firms to resort to frequent price cuts in order to wrest market share from competitors. Key to the long-term viability of price cuts is the ability of the firm to achieve cost savings through, e.g., outsourcing to providers from emerging markets (Kalaigianan & Varadarajan, 2012). This suggests that outsourcing to providers from emerging markets may be pursued to a greater extent in industries characterized by higher levels of competitive intensity. At the country level, we include a customer satisfaction index, which captures the corporate emphasis on customer satisfaction in the outsourcing firm's home country. Selecting an emerging market as an outsourcing location is a less likely strategy in countries where firms emphasize customer satisfaction, since such a strategy may negatively affect consumer perceptions (Thelen et al., 2010). Finally, we include a yearly trend variable to capture the increasing popularity of customer-support outsourcing to providers from emerging markets.

We then compute the inverse Mills ratio. In the second step, we add the inverse Mills ratio to the right hand side of Eq. (4) to control for

\(^5\) The continent clusters are Asia, Australia, Europe, and North America.

\(^7\) We specify the selection model as follows: \(Emerge = \alpha_0 + \alpha_1 PersCont + \alpha_2 KnowEmb + \alpha_3 Mktnt + \alpha_4 Size + \alpha_5 Profit + \alpha_6 Leverage + \alpha_7 IndCon + \alpha_8 SatEmph + \alpha_9 Trend + \epsilon\), where Profit is net income divided by sales in the year before the outsourcing announcement, Leverage is the ratio of long-term debt to total assets in the year before the outsourcing announcement, IndCon is a measure of industry concentration equal to the sum of the market shares of the four largest firms in the industry (low values indicate high competitiveness), SatEmph is a customer satisfaction index which measures the extent to which customer satisfaction is emphasized by companies in the outsourcing firm's home country (taken from the World Competitiveness Yearbook), and Trend is a yearly trend variable. The results show that outsourcing to providers from emerging markets is more likely when the outsourcing firm's past profitability is higher (\(\alpha_9 = .04, p < .01\)), when its leverage is lower (\(\alpha_6 = -2.66, p < .01\)), in industries characterized by higher levels of competitive intensity (\(\alpha_7 = -.92, p < .05\)), and in countries where the corporate emphasis on customer satisfaction is lower (\(\alpha_5 = -18, p < .10\)). We further find that a firm's probability to outsource to a provider from an emerging market increases over time (\(\alpha_9 = .12, p < .01\)), indicating the increasing popularity of this practice. The model shows acceptable fit with a pseudo R² of 17%. Interestingly, none of the focal variables in our study (PersCont, KnowEmb, Mktnt, Size) were significantly related to a firm's propensity to outsource to a provider from an emerging market (\(p > .10\)).
potential selection bias. An insignificant inverse Mills ratio indicates that selection bias is not likely to be a concern.

3.3. Sample

We gathered customer-support outsourcing announcements through extensive searches in the Lexis Nexis, Factiva, and SDC Platinum databases over the 1993–2007 period. This search resulted in an initial sample of 167 outsourcing announcements for firms from established economies that outsource customer-support activities to either established or emerging economies. Elimination of firms that were not publicly traded reduced the sample to 114 announcements. We further removed eleven announcements for firms for which stock price information was missing around the event day. To minimize the presence of confounding effects that might have extraneous influences on stock prices, we deleted 16 more announcements that included information about other important firm events (e.g., firm sales, earnings, CEO appointment) or if another announcement concerning the same firm appeared within a four-day window around the announcement.

The final sample of 87 announcements spans 21 industries. All outsourcing firms come from established economies, with the majority of outsourcing firms coming from the United States (40%), the United Kingdom (20%), and the Netherlands (7%). Most outsourcing firms are active in the communications (Standard Industrial Classification (SIC) code 48), business services (SIC code 73), and industrial machinery and equipment (SIC code 35) industries. The outsourcing providers come from a wide variety of countries, namely Australia, Belgium, Canada, Estonia, Germany, India, Indonesia, Ireland, Japan, Malaysia, Mexico, the Netherlands, Pakistan, the Philippines, Singapore, South Africa, Spain, Sweden, Taiwan, the U.K., and the U.S. One third of the outsourcing firms in our sample outsource to a service provider from an emerging market (with India being the most popular outsourcing location), whereas two third opt for an outsourcing provider from an established economy. Of the latter, 67% outsource to a service provider that is located in their home country (i.e., domestic outsourcing).

3.4. Operationalization

3.4.1. Financial measures

We obtained data on stock prices and market-wide indices from the Center for Research on Security Prices (CRSP) and from Datastream. To capture global market movements, we use the Financial Times Stock Exchange (FTSE) World Index. To account for foreign currency exchange rate fluctuations, we use the exchange rate between the U.S. dollar and the local currencies (cf. Park, 2004).

3.4.2. Emerging market versus established economy

We use a dummy variable that equals one when the outsourcing firm selects an outsourcing provider in an emerging market, and that is zero when it selects a provider in an established economy. Following Burgess and Steenkamp (2006, p. 339), we define an emerging market as “a country in which PPP-adjusted GDP per capita, converted to U.S. dollar and smoothed for three-year currency fluctuations, is equal to or less than the highest ranked country classified as ‘middle income’ by the World Bank.”

3.4.3. Nature of the outsourced customer-support service

We content-analyzed the outsourcing announcements to identify whether the outsourcing arrangement involved personal customer contact (e.g., telephone support) or impersonal customer contact (e.g., e-mail support). In a similar vein, we identified whether the outsourcing provider required highly embedded, complex knowledge for delivering customer support (e.g., technical customer support) or whether the outsourced customer-support processes were simple services characterized by low knowledge embeddedness (e.g., reservations). Two coders were introduced to the theoretical concepts “personal customer contact” and “high knowledge embeddedness,” and how they differ from “impersonal customer contact” and “low knowledge embeddedness.” Subsequently, the coders independently evaluated each outsourcing announcement on the basis of the construct definitions. Inter-coder agreement was over 90%. Differences between the coders were reconciled through in-depth discussion between the coders. We operationalized personal customer contact as a dummy variable that equals one when the outsourcing arrangement involves personal customer contact and zero otherwise. Similarly, we use a dummy variable that equals one for customer support that is characterized by high knowledge embeddedness and zero otherwise.

3.4.4. Marketing resource intensity

We measure the outsourcing firm’s marketing resource intensity as its annual Selling, General and Administrative (SG&A) expenditures divided by total assets. Although the SG&A measure also contains items that are not strictly marketing expenses, it is the best publicly available measure of marketing spending (see, e.g., Dinner et al., 2009; Luo, 2008 for similar practice).

3.4.5. Firm size

Following Gielens et al. (2008), we measure the outsourcing firm’s size as its total sales, one year prior to the outsourcing announcement. We log-transform this measure to reduce skewness and to account for potential diminishing returns to scale (see, e.g., Dekimpe, François, Gopalakrishna, Lilien, & van den Bulte, 1997).

3.4.6. Control variables

Labor-cost savings are the primary reason for most companies to outsource customer support to emerging markets. We therefore control for labor-cost differences between the countries of the outsourcing firm and the outsourcing provider. We divide labor costs in the country of the outsourcing firm by labor costs in the country of the outsourcing provider in the year prior to the announcement. Higher scores reflect that the outsourcing firm takes advantage of the lower labor costs in the country of the outsourcing provider. Approximately 22% of the outsourcing firms in our sample opted for an outsourcing provider from an established economy that differed from their home country (as opposed to selecting a domestic provider). Prior research has shown that consumers believe that domestic workers are superior to foreign service providers in serving them (Thelen et al., 2011, p. 272). Moreover, consumers associate outsourcing providers with a foreign accent with lower service quality, competence, and credibility (Stringfellow et al., 2008; Thelen et al., 2011). To account for such an effect, we include a domestic-outsourcing dummy that equals one when firms outsource to providers from their home country, and is zero otherwise.

We further control for industry concentration, which is measured as the sum of the market shares of the largest four firms in the outsourcing firm’s industry (cf. Cleeren, van Heerde, & Dekimpe, 2013), to account for the fact that outsourcing strategies may more closely fit with the environmental demands of less concentrated (i.e. more competitive) industries (Schilling & Steensma, 2001).

[^9]: Nine announcements pertained to outsourcing arrangements that involved both personal and impersonal customer contact. We assessed the robustness of our findings by removing these nine cases from our sample. Results remained substantively the same.

[^10]: Consistent with this line of research, we thus assume that, e.g., an Indian provider will not outperform a French provider, even though English is an official language in India. Both Indian and French speakers have an English accent that is likely to negatively affect the ease of communication. Nevertheless, we further examined whether communication-style problems are less severe if the outsourcing firm and the provider share an official language. We added a dummy variable to Eq. (4) that equals one when the outsourcing firm and the outsourcing provider share the same official language (e.g., U.S. firm–Indian provider or U.S. firm–U.K. provider), and is zero otherwise (e.g., U.S. firm–French provider). In line with our expectations, our results remained substantively the same.
Table 2
Variables and data sources.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Measure</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shareholder value</td>
<td>Changes in stock prices over a four-day event window using standardized cumulative abnormal returns (CAAR)</td>
<td>CRSP &amp; Datastream</td>
</tr>
<tr>
<td>Emerging market</td>
<td>Dummy variable: outsourcing firm outsource customer support to an outsourcing provider located in an emerging market (1) versus an established economy (0)</td>
<td>World Bank</td>
</tr>
<tr>
<td>Personal customer contact</td>
<td>Dummy variable that equals one when the outsourced customer-support service involves personal customer contact; zero otherwise</td>
<td>Lexis Nexis, Factiva, &amp; SDC Platinum</td>
</tr>
<tr>
<td>Knowledge embeddedness</td>
<td>Dummy variable that equals one when the outsourced customer-support service entails complex solutions that require highly embedded knowledge; zero otherwise</td>
<td>Lexis Nexis, Factiva, &amp; SDC Platinum</td>
</tr>
<tr>
<td>Marketing resource intensity</td>
<td>The outsourcing firm’s Selling, General, and Administrative (SG&amp;A) expenditures divided by total assets</td>
<td>Compustat &amp; Annual reports</td>
</tr>
<tr>
<td>Firm size</td>
<td>Ratio of labor costs in the country of the outsourcing firm and labor costs in the country of the outsourcing provider</td>
<td>Compustat &amp; Annual reports</td>
</tr>
<tr>
<td>Labor–cost savings</td>
<td>Total sales (log-transformed) of the outsourcing firm</td>
<td>World Development Indicators</td>
</tr>
<tr>
<td>Domestic outsourcing</td>
<td>Dummy variable that equals one when the outsourcing firm outsource customer support domestically; zero otherwise</td>
<td>Lexis Nexis, Factiva, &amp; SDC Platinum</td>
</tr>
<tr>
<td>Profitability</td>
<td>Net income divided by sales in the year prior to the outsourcing announcement</td>
<td>Compustat &amp; Annual reports</td>
</tr>
<tr>
<td>Leverage</td>
<td>Ratio of long-term debt to total assets in the year prior to the outsourcing announcement</td>
<td>Compustat</td>
</tr>
<tr>
<td>Industry concentration</td>
<td>Sum of market shares of the four largest firms in the outsourcing firm’s industry</td>
<td>Compustat &amp; Annual reports</td>
</tr>
</tbody>
</table>

Following Luo (2007) and Raassens, Wuyts, and Geyskens (2012), we control for additional financial information of the outsourcing firm that may influence stock returns, namely its profitability (net income divided by sales in the year before the outsourcing announcement) and leverage (ratio of long-term debt to total assets in the year before the outsourcing announcement).

To control for unobserved heterogeneity across industries and years, we add dummy variables for each industry and year to Eq. (4). Instead of retaining all industry and year dummies, which would lead to unstable results and multicollinearity, we estimate a trimmed model in which only the dummies that are significant at p < .10 are retained (cf. Anderson & Weitz, 1989).

A summary description of all measures, including the diverse data sources used, can be found in Table 2. Table 3 reports descriptive statistics (in original metrics) and correlations. Bivariate correlations exceeding .8 (Judge, Hill, Griffiths, Lutkepohl, & Lee, 1988) and variance inflation factors exceeding 10 (Mason & Perreault, 1991) indicate potential multicollinearity problems. All correlations and variance inflation factors are well below these critical values. Hence, multicollinearity is not a concern.

4. Results

Of all windows surrounding the event day, the one from −1 to +2 shows the most significant CAAR: CAAR[−1, +2] = −.15% (p < .05). This implies that, on average, the customer–support outsourcing announcement led to a decrease of .15% of shareholder value. Although customer–support outsourcing is, on average, evaluated negatively by the market, the performance implications differ substantially across outsourcing firms. Whereas 49% of the outsourcing firms show a negative CAAR, 51% were evaluated positively by investors. To explain this cross-sectional variation, we estimated Eq. (4).

Table 3
Descriptive statistics and correlation matrix.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>s.d.</th>
<th>Min.</th>
<th>Max.</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
<th>8.</th>
<th>9.</th>
<th>10.</th>
<th>11.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standardized CAAR[−1, +2]</td>
<td>.18</td>
<td>2.01</td>
<td>−4.62</td>
<td>5.66</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emerging market</td>
<td>.33</td>
<td>.47</td>
<td>.00</td>
<td>1.00</td>
<td>.09</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal customer contact</td>
<td>.67</td>
<td>.47</td>
<td>.00</td>
<td>1.00</td>
<td>.05</td>
<td>.03</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge embeddedness</td>
<td>.34</td>
<td>.48</td>
<td>.00</td>
<td>1.00</td>
<td>−.15</td>
<td>−.10</td>
<td>−1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marketing resource intensity</td>
<td>.26</td>
<td>.33</td>
<td>.00</td>
<td>1.87</td>
<td>−.07</td>
<td>−.03</td>
<td>−.00</td>
<td>.30</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firm size (in millions $)</td>
<td>15,682</td>
<td>25,069</td>
<td>4.64</td>
<td>155,445</td>
<td>.07</td>
<td>.10</td>
<td>.18</td>
<td>.14</td>
<td>.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor–cost savings</td>
<td>7.28</td>
<td>11.64</td>
<td>.64</td>
<td>72.68</td>
<td>.02</td>
<td>−.64</td>
<td>.06</td>
<td>.07</td>
<td>−.01</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domestic outsourcing</td>
<td>.45</td>
<td>.50</td>
<td>.00</td>
<td>1.00</td>
<td>.02</td>
<td>−.64</td>
<td>.10</td>
<td>.08</td>
<td>.18</td>
<td>−.30</td>
<td>−.49</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profitability</td>
<td>−.97</td>
<td>5.64</td>
<td>−46.30</td>
<td>.78</td>
<td>.04</td>
<td>.11</td>
<td>.09</td>
<td>−.20</td>
<td>−.18</td>
<td>.40</td>
<td>.09</td>
<td>.19</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leverage</td>
<td>−.14</td>
<td>.16</td>
<td>.00</td>
<td>.63</td>
<td>−.03</td>
<td>−.24</td>
<td>.15</td>
<td>−.15</td>
<td>−.28</td>
<td>.16</td>
<td>−.11</td>
<td>.09</td>
<td>.04</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Industry concentration</td>
<td>.61</td>
<td>.25</td>
<td>.19</td>
<td>1.00</td>
<td>.02</td>
<td>−.15</td>
<td>.07</td>
<td>.17</td>
<td>.16</td>
<td>.06</td>
<td>−.11</td>
<td>.03</td>
<td>−.01</td>
<td>−.13</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 4 presents the results. We find that the effect on shareholder value of outsourcing customer support is, on average, more favorable for outsourcing to providers located in emerging markets as opposed to established economies ($β_1 = 2.61, p < .05$).

Table 4
Drivers of the stock market reaction to outsourcing customer support.

<table>
<thead>
<tr>
<th>Hypothesized sign</th>
<th>b*</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>−.95</td>
<td>−.88</td>
</tr>
<tr>
<td>Emerging (1) vs. established market (0)</td>
<td>2.61**</td>
<td>4.93</td>
</tr>
<tr>
<td>Customer-support characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal customer contact</td>
<td>.77</td>
<td>1.96</td>
</tr>
<tr>
<td>Knowledge embeddedness</td>
<td>−.33</td>
<td>.44</td>
</tr>
<tr>
<td>Personal customer contact + Emerging</td>
<td>−1.90***</td>
<td>−3.53</td>
</tr>
<tr>
<td>Knowledge embeddedness + Emerging</td>
<td>−2.49***</td>
<td>−2.39</td>
</tr>
<tr>
<td>Firm characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marketing resource intensity</td>
<td>−1.33</td>
<td>−1.96</td>
</tr>
<tr>
<td>Firm size</td>
<td>−.22</td>
<td>−1.48</td>
</tr>
<tr>
<td>Marketing resource intensity + Emerging</td>
<td>.322**</td>
<td>3.38</td>
</tr>
<tr>
<td>Firm size + Emerging</td>
<td>.41**</td>
<td>2.14</td>
</tr>
<tr>
<td>Control variablesb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor–cost savings</td>
<td>−.00</td>
<td>−.15</td>
</tr>
<tr>
<td>Domestic outsourcing (1 = yes, 0 = no)</td>
<td>.28</td>
<td>.32</td>
</tr>
<tr>
<td>Profitability</td>
<td>−.03</td>
<td>−.78</td>
</tr>
<tr>
<td>Leverage</td>
<td>2.82</td>
<td>1.02</td>
</tr>
<tr>
<td>Industry concentration</td>
<td>2.43**</td>
<td>3.35</td>
</tr>
<tr>
<td>λ (inverse Mills ratio)</td>
<td>−1.52</td>
<td>−1.10</td>
</tr>
</tbody>
</table>

* p < .10.
** p < .05.
*** p < .01.

a We use one-sided tests for hypothesized effects, two-sided tests for non-hypothesized effects.

b For simplicity of presentation, the results for the industry and year dummies are not reported in the table.
As hypothesized \( H1 \), we find a negative interaction effect between customer-support outsourcing to emerging (versus established) markets and personal customer contact \( H1: \beta_4 = -1.90, p < .05 \). Thus, the performance advantage of outsourcing to emerging markets is lower when the outsourced customer support involves personal customer contact. Also \( H2 \) is supported. The performance advantage of outsourcing to emerging (versus established) markets is lower when the outsourced customer-support activity is characterized by high knowledge embeddedness \( \beta_5 = -2.49, p < .05 \).

As hypothesized in \( H3 \), we find a positive significant interaction effect between customer-support outsourcing to emerging (versus established) economies and the outsourcing firm’s marketing resource intensity \( \beta_6 = 3.22, p < .05 \). \( H4 \) proposes a positive interaction effect between customer-support outsourcing to emerging (versus established) economies and the outsourcing firm’s size. Also this hypothesis is supported \( \beta_7 = .41, p < .10 \). The coefficient of the inverse Mills ratio was not significant \( \lambda = -1.52, p > .10 \). Thus, endogeneity was not a major concern in our study.

5. Robustness checks

To check the robustness of our findings, we perform three additional analyses.

5.1. Alternative event window

Our measure of performance is the cumulative abnormal return over the 4-day event window \([-1,2]\). We validate our results across the alternative window \([-1,1]\). We opt for this window because it is at the same time short enough to benefit from increased power of the test statistic (McWilliams & Siegel, 1997), and long enough to deal with the lack of synchronism in stock-market trading hours between countries \(5–6\) h difference between American and European countries and between European and most Asian countries; Park, 2004). Our results are robust for this alternative event window. Specifically, we replicate the positive and significant main effect on shareholder value of outsourcing to providers in emerging (rather than established) economies \( p < .05 \). Further, all interaction effects have the expected sign and are significant \( p < .10 \).

5.2. Long-run correction

To test the efficient market assumption, we check whether the initial evaluation was not just a short-run over- or under-reaction that was corrected in the longer-run (Fama, 1998). We calculate 3-months, 6-months, 1-year, and 2-year long-term effects using the buy-and-hold abnormal returns (BHAR) and Libbott’s returns across time and securities (IRATS) models. We find no effect on long-term abnormal returns \( p > .10 \), suggesting that the stock market is reasonably efficient. Thus, the abnormal returns to outsourcing customer support occur in the short-term window and there are no corrections in the long-run.\(^{10} \)

5.3. Validation of theoretical arguments

The arguments underlying our hypotheses are based on the markedly different cultural and regulative systems of emerging markets and established economies. In particular, the theory underlying \( H1 \) and \( H2 \) pertains to the communication-style problems that are inherent to the cultural system of emergent markets, while the theory behind \( H3 \) and \( H4 \) relates to emerging markets’ regulative system. To validate our line of reasoning, we re-estimate our model after replacing our focal variable \( Emerge \), which captures the selection of an emerging market rather than an established economy as outsourcing location] with two continuous measures: (i) a cultural-distance measure, to capture communication-style problems, and (ii) a rule-of-law measure, to capture the regulatory framework in the country of the outsourcing provider. \(^{11} \)

Table 5 presents the results. We find that the interactions

\(^{10} \) Our dependent variable is the short-term cumulative abnormal return accruing from the outsourcing announcement to the outsourcing firm. Although this measure materializes in the short term (consistent with the efficient market hypothesis that a firm’s stock price immediately reflects all new information, conceptually this measure reflects the stock markets’ best estimate of the change in the long-term value of the firm. Significant long-term abnormal returns, as computed by the BHAR or IRATS methodology, reflect that the stock market is not efficient, but rather over- or under-reacted to the announcement. They do not connote long-term firm performance.

\(^{11} \) Cultural distance is operationalized as a composite index, using the cultural dimensions (power distance, uncertainty avoidance, individualism, and masculinity) developed by Hofstede (2001) and the operationalization by Kagdet and Singh (1988). Rule-of-law is operationalized as the extent to which agents have confidence in and abide by the rules of society (i.e. the quality of contract enforcement, property rights, the police, and the courts, and the likelihood of crime and violence). The measure is taken from Kaufmann, Kraay, and Mastruzzi (2008). Both variables are highly correlated with \( Emerge \), \( r = .70 \) for cultural distance and \( r = -.33 \) for rule-of-law) and with each other \( r = -.05 \).
between personal customer contact and cultural distance ($\beta = -1.15, p < .05$) and between knowledge embeddedness and cultural distance ($\beta = -52, p < .10$) are significant and in the expected direction. As expected, we also find significant negative interactions between marketing resource intensity and rule-of-law ($\beta = -1.78, p < .05$) and between firm size and rule-of-law ($\beta = -34, p < .05$). Hence, this analysis corroborates the underlying developed arguments.

6. Discussion

Customer-support outsourcing has received bad press, featuring alarming headlines such as “Outsourcing’s Impact on Customer Satisfaction: It’s Not Good.” “Beware: Your Customers Oppose Outsourcing,” and “Losing Money by Spending Less.” Despite the bad press, customer-support outsourcing is still on the rise, especially to service providers from emerging markets. The primary reason for firms to embrace a customer-support outsourcing strategy is to obtain cost savings. However, approximately 50% of the outsourcing firms in our sample experience negative stock market reactions. Thus, whereas outsourcing customer support works for some firms, it is not a one-size-fits-all strategy. Our focus is on explaining what makes customer-support outsourcing to providers from emerging versus established markets more versus less successful. Below, we discuss the theoretical and the managerial contributions of our research.

6.1. Theoretical implications

6.1.1. The moderating role of customer-support characteristics

Extant research (e.g., Bharadwaj & Roggeveen, 2008; Roggeveen et al., 2007) has repeatedly shown that the geographic location of the outsourcing provider influences consumer expectations and satisfaction, and that consumers prefer service support services from their own country. Based on this work, one might be tempted to conclude that customer support should never be outsourced offshore—and certainly not to emerging markets. We show that the logic is more intricate. While, on average, outsourcing to providers from emerging markets enhances shareholder value more than outsourcing to established economies, there is substantial variation depending on the nature of the outsourced customer-support service and the nature of the outsourcing firm.

To fully understand the performance implications of customer-support outsourcing, one should acknowledge the differences between outsourced customer-support services rather than generalize across them, which has been the convention in extant research. We find that outsourcing customer support to emerging markets is less beneficial for support services that are characterized by (1) direct, real-time interaction between the customer and the service employee (as opposed to impersonal contact), and/or (2) highly embedded knowledge (as opposed to more superficial knowledge).

6.1.2. The triadic nature of customer-support outsourcing

Our findings contribute to recent endeavors, primarily in the area of supply chain management, to understand service triads (e.g., Gunawardane, 2012; Van der Valk & Van Iwaarden, 2011). This emerging stream of literature builds on earlier conceptualizations by both operations management and marketing scholars of the triadic nature of exchange (e.g., Choi & Wu, 2009; Wuyts, Streemersch, Van den Bulte, & Franses, 2004). Service triads consist of three actors (the client firm, the service provider, and the end customer) and the ties that bind them (e.g., due to an outsourcing agreement). Outsourcing to providers from emerging markets increases communication difficulties in two ties: (1) the tie between the service provider’s employees and the customer, and (2) the tie between the outsourcing firm’s employees and the service provider’s employees. On the one hand, when service employees in emerging markets engage in direct, real-time interaction with customers in established economies, customers can become irritated and engage in negative behaviors toward the outsourcing firm.

On the other hand, when the outsourced service entails highly embedded knowledge, communication-style differences between employees of the service provider and employees of the outsourcing firm can complicate effective knowledge transfer. Whereas the extent marketing literature on customer-support outsourcing has extensively studied the tie between the service provider and the customer, the tie between the outsourcing firm and the service provider has received much less attention. Our results corroborate the need to acknowledge the triadic nature of outsourcing.

6.1.3. The moderating role of firm characteristics

The performance potential of outsourcing customer support to emerging markets as opposed to established economies also depends on the nature of the outsourcing firm. First, larger firms benefit more from outsourcing customer-support services to emerging markets than smaller firms, presumably because they can credibly threaten to switch to other providers and can select better partners through their preferential relationship with local governments. Second, outsourcing customer support to emerging markets leads to a higher performance potential for firms with a higher marketing resource intensity than for firms with a lower marketing resource intensity. This finding, along with the validation of a high marketing resource intensity can help firms hedge against a weak regulatory system, contributes to earlier efforts to understand the role of marketing (e.g., Kumar & Shah, 2009; Moorman & Rust, 1999; Verhoef & Leeflang, 2009). Drawing on Day’s (1994) seminal work, we bring back in marketing’s role in terms of environmental scanning and we argue that firms with a strong marketing focus are better able to identify appropriate exchange partners. As a consequence, these firms can better deal with the partner-related risk that is associated with weak regulatory environments. Conversely, our results suggest that a strong legal environment substitutes for some of the benefits offered by a strong marketing focus. This is in line with the results of Wu (2013), who shows that marketing capabilities have a weaker effect on firm performance in countries characterized by strong legal systems.

6.2. Managerial guidelines

To arrive at a better understanding of the impact of customer-support and firm characteristics on performance, we conduct a “what-if” analysis. We use the estimates from Table 4 to calculate the performance implications of outsourcing customer-support services characterized by personal (as opposed to impersonal) customer contact and requiring highly embedded (as opposed to more superficial) knowledge for small versus large-sized firms that are low versus high on marketing resource intensity. Table 6 presents the results.

We first turn to the left upper quadrant, i.e. the case where the outsourced customer-support service is personal and knowledge embeddedness is low. Irrespective of a firm’s size and marketing resource intensity, outsourcing to a provider from an emerging market has more positive performance implications than outsourcing to a provider from an established economy. The difference in performance implications between emerging and established markets is least pronounced for small firms with a low marketing resource intensity, but even these firms experience shareholder value increases that are approximately 40% higher when outsourcing to providers from emerging (CAR = 2.61) rather than established (CAR = 1.86) economies.

The picture reverses when the outsourced customer-support service is personal and knowledge embeddedness is high (right lower quadrant). In three out of the four cells, the shareholder value implications of outsourcing such services are higher for providers from established rather than emerging economies. For small firms with a low marketing
resource intensity, outsourcing personal and specialized customer support to providers from emerging markets even becomes value-destroying, as attested by the CAR of −.69. Only for large firms with a high marketing resource intensity, providers from emerging markets should be preferred over providers from established economies, although the performance difference is relatively small in magnitude (CAR = 1.37 vs. 1.05).

The other two quadrants show more variation. When outsourcing \textit{personal customer-support services that are low on knowledge embeddedness} (right upper quadrant), firms are generally better off selecting an outsourcing provider from an emerging market. Only small firms with a low marketing resource intensity would do better to outsource to providers from established economies (CAR = 2.63) rather than emerging markets (CAR = 1.48). When outsourcing \textit{personal customer-support services that are high in knowledge embeddedness} (left lower quadrant), the same conclusion can be drawn. Large firms are better off selecting an outsourcing provider from an emerging market (regardless of their marketing resource intensity). Small firms with a high marketing resource intensity are also better off selecting an outsourcing provider from an emerging market, but small firms with a low marketing resource intensity experience more favorable performance implications by selecting an outsourcing provider from an established economy.

In sum, when considering the different quadrants in Table 6 jointly, large firms always benefit more from outsourcing customer support to emerging markets rather than established economies, except when their marketing resource intensity is low and the service is both personal and high in knowledge embeddedness. For small firms, the picture is more intricate. In case of high knowledge embeddedness, they are better off selecting a provider from an established economy except when their marketing resource intensity is high and the service is impersonal in nature. In contrast, when small firms outsource services that are low on knowledge embeddedness, they are better off selecting a provider from an emerging market, except when their marketing resource intensity is low and the service is personal in nature.

An interesting observation stemming from our endogeneity analysis is that when outsourcing customer support to emerging markets versus established economies, firms did not yet take the performance implications into account. Hence, it appears that firms do not yet behave optimally in terms of the performance that could be obtained. A possible explanation stems from the fact that outsourcing to emerging markets is a fairly recent phenomenon, while it takes time to develop optimal practices (Nelson & Winter, 1982). Firms have little or no own prior experiences to rely on, and are therefore still uncertain about the benefits and drawbacks of their outsourcing decisions. Hence, our study provides several valuable first insights for firms intending to outsource customer support in the near future.

6.3. Future research directions

We organize our agenda for future research along five trajectories: (1) using more fine-grained measures, (2) uncovering heterogeneity among emerging markets, (3) further characterizing outsourcing firms and outsourcing providers, (4) solving the paradox between outsourcing marketing activities while not eroding internal capabilities, and (5) examining the full outsourcing firm—service provider—customer triad.

First, our measure of marketing resources is coarse. Although SG&A is a frequently used measure of marketing spendings (see, e.g., Dutta, Narasimhan, & Rajiv, 1999; Mizik & Jacobson, 2007), it does not only capture the firm’s expenditures in sales force, advertising, and promotional activities, but also covers general overhead and legal costs. Future research would benefit from a more fine-grained measure that isolates the marketing spending items.

Second, our study focuses on systematic differences between emerging and established economies. Fostered by current developments in the outsourcing market, where the most popular outsourcing locations (e.g., India) are overheating and losing ground to locations such as South Africa, Morocco, and Brazil, future research could study differences among emerging markets.

Third, our coverage of firm-related characteristics is only partial. Thus, future investigations could enhance our research by examining additional firm characteristics. In addition, a question that remains unanswered in this paper is to what extent the performance implications of outsourcing customer support depend on characteristics of the outsourcing provider. The outsourcing providers were often located in countries where firm-specific information is poorly documented. Hence, we were unable to test for characteristics of the outsourcing providers.

Table 6

<table>
<thead>
<tr>
<th>Performance potential (predicted CAR) of outsourcing customer support to emerging versus established economies as a function of customer-support and firm characteristics.</th>
<th>Impersonal customer-support service that is low in knowledge embeddedness</th>
<th>Personal customer-support service that is low in knowledge embeddedness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low marketing resource intensity</td>
<td>Established economy</td>
<td>Large firm</td>
</tr>
<tr>
<td>Emerging market</td>
<td>2.17</td>
<td>2.10</td>
</tr>
<tr>
<td>Established economy</td>
<td>1.40</td>
<td>.28</td>
</tr>
<tr>
<td>High marketing resource intensity</td>
<td>Established economy</td>
<td>Large firm</td>
</tr>
<tr>
<td>Emerging market</td>
<td>1.56</td>
<td>.50</td>
</tr>
<tr>
<td>Established economy</td>
<td>1.40</td>
<td>.28</td>
</tr>
</tbody>
</table>

The alternative with the highest performance potential is marked in bold.

To calculate predicted values, the indicator variables to measure the nature of the customer-support service are set to zero or one, while the continuous variables are fixed at their baseline level (i.e. zero for indicator control variables and the mean for continuous control variables).
Fourth, our finding that firms higher in marketing resource intensity benefit less from outsourcing customer support to established economies raises a broader strategic question related to the boundaries of the firm. How should firms balance nurturing internal marketing capabilities with outsourcing customer support? For any given marketing function, does nurturing internal capabilities substitute for outsourcing, or do internalized capabilities and outsourcing serve as complements? Whether firms should retain a certain degree of activities in-house to reap the benefits from outsourcing remains an issue for future research.

Fifth, outsourcing service delivery leads to a triadic relationship between the outsourcing firm, the service provider, and the customer. When customers are dissatisfied with service delivery by the service provider, this may have ramifications for the outsourcing firm—customer relationship. Customers may call upon the outsourcing firm to correct the outsourcing provider. Metters (2008) gives the example of a customer who complained to Dell about the “horrible accent an Indian call center worker had” and who threatened to forego future business with Dell (p. 205). More research is needed to understand this governance mechanism, which is known as two-step leverage in the network governance literature (Gargiulo, 1993; Wuyts, 2010).


Full Length Article

The effect of customer empowerment on adherence to expert advice

Nuno Camacho a,⁎, Martijn De Jong a, Stefan Stremersch a,b

a Erasmus School of Economics, Erasmus University, Rotterdam, The Netherlands
b IESE Business School, University of Navarra, Spain

A R T I C L E  I N F O
Article info
First received in September 20, 2012 and was under review for 6 months
Available online 24 April 2014
Area Editor: Russell S. Winer
Keywords:
Adherence to expert advice
Empowerment
International marketing research
Health marketing

A B S T R A C T
Customers often receive expert advice related to their health, finances, taxes or legal procedures, to name just a few. A noble stance taken by some is that experts should empower customers to make their own decisions. In this article, we distinguish informational from decisional empowerment and study whether empowerment leads customers to adhere more or less to expert advice. We empirically test our model by using a unique dataset involving 11,735 respondents in 17 countries on four continents. In the context of consumer adherence to doctors’ therapy advice (patient non-adherence to doctor advice may cost about $564 billion globally to the pharmaceutical industry every year), we find that decisional empowerment lowers adherence to expert advice. The effect of informational empowerment varies predictably across cultures and is only universally beneficial when initiated by the customer. These findings have important implications for professional service providers.

1. Introduction

Customers often rely on experts, such as accountants, consultants, lawyers and physicians to make complex decisions (Bove & Johnson, 2006). Expert advice decreases decision complexity (Brehmer & Hagañor, 1986) and may improve decision quality (Yaniv, 2004). There is a rich literature, in marketing and psychology, on customer–expert interactions. One stream of literature focuses on how experts use customers’ input and feedback to update their beliefs and decisions (e.g. Camacho, Donkers, & Stremersch, 2011; Narayanan & Manchanda, 2009). For instance, Camacho et al. (2011) show that, when learning about a new drug, physicians place more emphasis on feedback from patients who switch to alternative treatments than on feedback from patients who continue their therapy. A second stream of literature focuses on expert advice and customer adherence to such advice (Bonaccio & Dalal, 2006; Bowman, Heilman, & Seetharaman, 2004; Fitzsimons & Lehmann, 2004; Schwartz, Luce, & Ariely, 2011; Tost, Gino, & Larrick, 2012; Usta & Häubl, 2011). The present paper focuses on the effects of customer empowerment during an advising interaction on customer adherence to expert advice.

In a typical customer–expert interaction, a customer receives an advice from the expert and subsequently decides whether to adhere to such advice.1 A robust finding from this literature is that people do not sufficiently adhere to expert advice (Bonaccio & Dalal, 2006). The traditional view of customer–expert interactions is that the expert should choose a particular course of action on behalf of the customer (e.g. “I would advise you to do X”, see Bonaccio & Dalal, 2006, p.128), a decision-making style we call “paternalistic” (e.g. Charles, Gafni, & Whelan, 1999). For example, a paternalistic lawyer–client interaction proceeds with a client exposing a legal problem to her lawyer who then recommends a particular course of action to the client (Macfarlane, 2008). The lawyer then expects the client to follow her advice to maximize chances of successful litigation.

⁎ We thank the Marketing Science Institute for its financial support. We are also indebted to Bas Donkers, Benedict Dei Daert and Gerry Tellis for their insightful comments and suggestions. We also thank the participants at the 2010 Marketing Science and 2011 EMAC Conferences and the seminar participants at Católica Lisbon – School of Business and Economics, Groningen University and Tilburg University for all their valuable suggestions. The usual disclaimer applies.

1 We assume a setting where the customer seeks the advice of a single expert and that customer adherence to the expert’s advice improves decision quality for the customer. This assumption builds upon the advice-taking literature (Bonaccio & Dalal, 2000; Yaniv, 2004; Yaniv & Kleinberger, 2000).
Customer overconfidence may increase both unintentional and reasoned non-adherence. On the one hand, overconfident customers tend to listen less carefully to expert advice (Tost et al., 2012), which increases unintentional non-adherence. On the other hand, overconfident customers tend to egocentrically discount the expert’s advice (See et al., 2011; Yaniv, 2004), which increases reasoned non-adherence.

Using a multi-sample Bayesian structural equation model, we show that decisional empowerment is associated with higher unintentional and reasoned non-adherence to expert advice and that informational empowerment is only able to reduce unintentional and reasoned non-adherence when the customer explicitly requests the exchange of additional solution-relevant information. We empirically validate our expectations in the highly relevant domain of healthcare decisions. Consumer non-adherence to doctor advice contributes to disease progression and increased mortality rates, resulting in annual direct and indirect healthcare costs of at least $290 billion in the U.S. alone (New England Healthcare Institute, 2009) and lost revenue for pharmaceutical firms of $564 billion a year globally (Forissier & Firlik, 2012).2

Our sample includes 11,735 respondents in 17 countries on four continents. To the best of our knowledge, this is by far the largest and geographically most diverse test of the relationship between customer empowerment and adherence to date. Prior empirical research on the relationship between empowerment and adherence to expert advice has focused on the U.S. or a selected set of Western nations, while customers’ reaction to empowerment may be vastly different across cultures (Botti, Orfali, & Iyengar, 2009; Charles, Gafni, Whelan, & O’Brien, 2006).

We build upon Schwartz’s (1994) cultural values theory, to explain systematic cross-country differences in the relationship between customer empowerment and adherence to expert advice. Our analyses revealed that, in line with our expectations, culture matters. We find that culture moderates the effects of decisional empowerment and, to a lesser extent, of informational empowerment on non-adherence in systematic and predictable ways. These findings have important implications for marketers and policy makers.

2. Theoretical background: Customer empowerment and adherence to expert advice

The expert advice literature typically distinguishes between advice-taking (Bonaccio & Dalal, 2006; Yaniv, 2004). We first organize advice-giving styles according to customer empowerment. Next, we discuss advice-taking, which, in our context, is the customer’s decision to adhere or deviate from the expert’s advice.

2.1. Organizing advice-giving styles according to customer empowerment

Fig. 1 organizes different advice-giving styles, according to informational empowerment, through expert facilitation (x-axis) or customer initiative (y-axis), and decisional empowerment (the z-axis). Expert facilitation of informational empowerment happens when the expert proactively exchanges solution-relevant information with the customer (i.e. takes the initiative of sharing solution-relevant information even if it is not requested by the customer). Customer-initiated informational empowerment happens when the customer requests solution-relevant information from the expert. Under decisional empowerment, the customer retains autonomy over the decision, which is the opposite of decision delegation by the customer to the expert (see Usta & Häubl, 2011).

The advice-giving styles at the bottom of Fig. 1 are characterized by low decisional empowerment (i.e. choice delegation), while those at the top are characterized by high decisional empowerment (i.e. choice autonomy).
In the bottom left of the graph, we depict the traditional paternalistic model which is characterized by low decisional empowerment and by low informational empowerment (Charles et al., 1999). In a paternalistic model, the expert decides on behalf of the customer in a paternalistic manner and hence only needs to exchange the information needed to identify and understand the customer’s problem (diagnostic information). In informed delegation models, customers and experts also exchange solution-relevant information. Conditional on the information collected, the expert then applies her knowledge to choose an option that maximizes the customer’s utility (Phelps, 1992).

At the top of Fig. 1, we depict consumerist and informed autonomy models. In consumerist models (Coulter, 1999), the customer demands that the expert help her execute a self-chosen course of action and there is no exchange of solution-relevant information. Examples of consumerism include requests for a specific litigation strategy by clients to their lawyers (Macfarlane, 2008) and branded request by patients to their doctors (Venkataraman & Stremersch, 2007), a phenomenon that has steadily increased in recent years (Stremersch, Landsman, & Venkataraman, 2013). To the extent that the customer takes initiative in exchanging solution-relevant information during her interaction with the expert, consumerism can yield customer-driven informed autonomy (Charles et al., 1999). In the expert-driven informed autonomy model (Quill & Brody, 1996), the expert facilitates the exchange of solution-relevant information, but leaves the final choice of a course of action to the customer.

2.2. Advice-taking: Customer adherence to expert advice

We conceptualize adherence to expert advice as the propensity of a customer to follow an expert’s advice (Bonaccio & Dalal, 2006; DiMatteo et al., 1993). Adherence to expert advice requires an effortful commitment of the customer to implement the behaviors recommended by the expert during the advising interaction. If customers have difficulty to understand or recall some of the information transmitted by the expert (e.g., the different steps a tax advisor recommended his client to minimize her tax payments), they may unintentionally non-adhere to the advice. If customers do not accept and deliberately deviate from the expert’s advice (and rely more on their own opinion than on the expert’s opinion), we speak of reasoned non-adherence (Bonaccio & Dalal, 2006; Yaniv & Kleinberger, 2000).

3. Hypotheses development

In developing hypotheses about the effects of customer empowerment on adherence to expert advice, we rely on two key psychological mechanisms: (1) dual models of information processing (Chaiken, 1980; Petty & Cacioppo, 1986) and (2) customer overconfidence (See et al., 2011; Tost et al., 2012; Yaniv & Kleinberger, 2000).

Dual models of information processing, such as the heuristic systematic model (HSM; Chaiken, 1980; Chaiken, Liberman, & Eagly, 1989) and the elaboration likelihood model (ELM; Petty & Cacioppo, 1986) posit that customers possibly engage in two modes of information processing, which involve different levels of thought and cognitive effort. Heuristic (or peripheral) processing is relatively effortless and quick while systematic (or central) processing requires customers to devote more cognitive resources to process information. A good example, in a healthcare context, is provided by Stegina and Occhipinti (2004) who show, for patients with prostate cancer, that customers may either use an expert opinion heuristic (e.g., “experts can be trusted”, p.574) or more systematic information processing strategies (e.g. weighing all pros and cons of different recommended options). For these reasons, dual-process models have special relevance for the effects of informational empowerment on unintentional non-adherence.

Recent research in social psychology suggests that empowerment may lead people to feel more powerful in a relationship and become overconfident about their abilities (See et al., 2011). Overconfident customers tend to overweight their own knowledge and opinions and therefore: (i) listen less carefully to expert advice (Tost et al., 2012)
and (ii) egocentrically discount expert advice (Bonaccio & Dalal, 2006; See et al., 2011; Yaniv, 2004). Moreover, when given power in a certain decision task, people tend to generalize their overconfidence to tasks outside the original scope of empowerment (Weitlauf, Cervone, Smith, & Wright, 2001). Hence, customer overconfidence has special relevance for the effects of decisional empowerment on non-adherence and for the effects of informational empowerment on reasoned non-adherence.

3.1. Expert facilitation of informational empowerment and customer non-adherence

Expert facilitation occurs when an expert proactively exchanges solution-relevant information with the customer during an advising interaction (e.g. a doctor asks a child whether she likes strawberries or cherries to decide on a drug’s flavor to prescribe, or a lawyer discusses with a client which expert witness to appoint in a patent litigation case). Experts often exchange unrequested solution-relevant information with customers in order to increase the customer’s involvement and responsibility in a given decision-making task (Epstein et al., 2004). Dual-process models predict that elevated responsibility increases task importance and thus motivates customers to use systematic, rather than heuristic, information processing (Bohner, Moskovitz, & Chaiken, 1995; Chaiken, 1980). However, systematic processing of unrequested pieces of information may increase customers’ cognitive and emotional burden and eventually obscure other relevant pieces of information (Epstein, Korones, & Quill, 2010).

Thus, when compared with a paternalistic model, expert facilitation of informational empowerment requires the customer to systematically process additional solution-relevant information. Such additional information will compete, in the customer’s memory, with other key pieces of information in the expert’s advice (e.g., dosing instructions in a patient–physician interaction or advice on specific litigation steps in a lawyer–client interaction), making the latter less salient and the advice harder to recall⁴ (Raaijmakers & Shiffrin, 1992), as compared to a paternalistic interaction. Forgetting, in turn, is one of the key reasons why customers do not adhere to expert advice (Osterberg & Blaschke, 2005). Thus, we expect that:

**H1.** Expert facilitation of informational empowerment increases unintentional non-adherence.

Customers often suffer, in their relationship with advisors, from “egocentric bias”, i.e. from a tendency to overweight their own opinion and egocentrically discount the expert’s advice (Bonaccio & Dalal, 2006; Yaniv & Kleinberger, 2000). This means that even when a customer accepts that the expert’s advice is correct, she may still depart from this advice and maintain her own prior attitudes and beliefs, resulting in reasoned non-adherence (Bonaccio & Dalal, 2006). Expert facilitation of informational empowerment may increase this tendency. When compared with a paternalistic model, expert facilitation of informational empowerment may elevate customers’ perceived power in the customer–expert relationship, i.e. the belief in their own ability to decide and control the problem being discussed (Tost et al., 2012). Customers with an elevated perceived power tend to become overconfident, which leads them to place more weight in their own beliefs and less weight in the expert’s advice (Bonaccio & Dalal, 2006; See et al., 2011; Yaniv, 2004). Therefore, expert facilitation of informational empowerment may trigger customers to egocentrically discount the expert’ advice more than a paternalistic customer–expert interaction. Therefore, we hypothesize the following:

**H2.** Expert facilitation of informational empowerment increases reasoned non-adherence.

3.2. Customer-initiated informational empowerment and customer non-adherence

Customer-initiated informational empowerment results in the discussion of solution-relevant information that the customer finds self-relevant and meaningful. Prior research in dual-process models shows that self-relevance triggers systematic information processing (Chaiken, 1980). Systematic processing of self-relevant information should increase customers’ motivation to carefully listen to the advice (Ryan & Deci, 2000), which, in turn, facilitates understanding and recall of the information exchanged. For instance, Kreuter, Clark, Oswald, and Bull (1999) show that cognitive elaboration focused on self-relevant information facilitates understanding and future recall of health-related advice. Similarly, Brug, Steenhuis, Van Assema, and de Vries (1996) find that people who receive nutrition advice customized to their personal dietary behavior perceive such advice as self-relevant and adhere more to advice than people who receive non-tailored advice. In line with this logic, Abele and Gendolla (2007) show that active exercisers process health information focusing on physical exercise more deeply, and recall it better, than non-active exercisers. Thus, we expect that:

**H3.** Customer-initiated informational empowerment decreases unintentional non-adherence.

Customer-initiated informational empowerment may also affect reasoned non-adherence. When compared with a paternalistic model, customer-initiated informational empowerment may affect the distribution of perceived power between the customer and the expert in different ways. The effect thereof on reasoned non-adherence is unclear. On the one hand, it may be possible that the customer gains power in the customer–expert relationship. This happens if the customer discovers, in the expert’s response to her request for solution-relevant information, evidence that contradicts the expert’s advice (Chaiken et al., 1989). Contradictory information enables the customer to challenge the validity of the expert advice, which may increase the customer’s perceived power relative to the expert.

On the other hand, it may also be conceivable that the expert gains power in the customer–expert relationship. For instance, the expert may push back the customer’s initiative and refuse to discuss solution-relevant information. When compared with a paternalistic interaction, an expert’s refusal to respond to a customer’s requests for additional information avoids the increase in perceived power, and subsequent customer overconfidence, discussed above (Izraeli & Jick, 1986). Alternatively, the expert may, through skillfully answering the questions posed by the customer, increase her expert status and undermine customer overconfidence.

Hence, when compared to a paternalistic interaction, customer-initiated informational empowerment may increase, or decrease, the customer’s tendency to egocentrically discount the expert’s advice (Yaniv & Kleinberger, 2000). Given these conflicting expectations, the ultimate effect of customer-initiated informational empowerment on reasoned non-adherence will depend on which of these two forces dominates and is, thus, an empirical question.

3.3. Decisional empowerment and therapy non-adherence

Decisional empowerment may increase unintentional non-adherence in two main ways. First, decisional empowerment may trigger customer overconfidence and worse information processing. Decisional empowerment allows customers to feel in control of their decisions,
and increases their power in the customer–expert relationship (Botti & McGill, 2011). As discussed above, power may trigger overconfidence (see et al., 2011). Hence, when compared with a paternalistic model, decisional empowerment should lead customers to overestimate the accuracy of their beliefs and opinions, which leads them to listen and process the expert advice less carefully (Tost et al., 2012). Less careful processing of the advice increases the likelihood that the customer forgets key components of the advice.

Second, decisional empowerment increases the customer’s responsibility in decision-making, potentially magnifying the emotional and cognitive costs of the decision task (Botti & McGill, 2011; Botti et al., 2009). These effects may increase customer anxiety (Botti et al., 2009), which, in turn, has been shown to impair information processing (Sengupta & Johar, 2001). Consequently, decisional empowerment impairs the quality of the customer–expert communication and reduces the salience of the expert’s advice making it harder to recall later. We thus hypothesize:

**H4.** Decisional empowerment increases unintentional non-adherence.

Decisional empowerment may also increase the likelihood of reasoned non-adherence. When compared with expert facilitation of informational empowerment, decisional empowerment represents a stronger departure from the traditional paternalistic customer–expert relationship (Charles et al., 1999; Quill & Brody, 1996). In addition, decisional empowerment entails patient participation in the decision-making without necessarily allowing the customer to learn more about the problem under discussion. Hence, as discussed above, decisional empowerment may elevate customer power and trigger overconfidence, which should lead customers to place less weight on the expert’s opinion and egocentrically discount the expert advice (See et al., 2011; Yaniv & Kleinberger, 2000).

In addition, overconfident customers tend to generalize their self-efficacy perceptions from a focal decision domain to decision domains outside the original scope of empowerment (Weitlauf et al., 2001). Accordingly, decisional empowerment during an advising interaction (e.g. participating in the choice of one out of several alternative courses of action) may lead customers to become overconfident about their capacity to decide when to alter or stop their adherence to expert advice, increasing reasoned non-adherence. In the words of Bowman et al. (2004), in the context of physicians empowering patients to make their own treatment choices, the “perception of empowerment and control should persist such that the consumer also believes that he or she is capable of changing dosage or stopping usage altogether without physician consultation” (p. 325). Therefore, we hypothesize that:

**H5.** Decisional empowerment increases reasoned non-adherence.

### 3.4. Cultural effects

Behavioral responses to customer empowerment may be vastly different across different national cultures (Charles et al., 2006). In particular, we expect national-cultural values to shape expectations about the role of experts and to trigger positive or negative social reinforcement mechanisms that moderate the effects of customer empowerment on non-adherence. This fits the tradition in international marketing of considering national-cultural values as moderators of customer behavior (Burgess & Steenkamp, 2006; Steenkamp & De Jong, 2010; Steenkamp & Geyskens, in press; Stermersch & Lemmens, 2009; Stermersch & Tellis, 2004; Tellis, Stermersch, & Yin, 2003; Van den Bulte & Stermersch, 2004; van Everdingen, Fok, & Stermersch, 2009).

We adopt Schwartz’s (1994) framework of national-cultural values, instead of the alternative frameworks of Hofstede, Inglehart and Baker, or Triandis (see Vinken, Soeters, & Ester, 2004, for an overview), for three key reasons. First, Schwartz derived his cultural dimensions from his individual-level theory of human value priorities (Schwartz, 1992), which is one of the most widely validated theories in social sciences (Schwartz et al., 2001). For this reason, Schwartz’s (1994) cultural framework is conceptually the most pure among the existing theories of national-cultural values (Bond et al., 2004; Burgess & Steenkamp, 2006).

Second, this framework is robust in terms of its measurement properties. The different value dimensions in this framework form an integrated and interdependent system, in contrast to other frameworks in which cultural dimensions are orthogonal to each other (e.g. Hofstede, 2001; Inglehart & Baker, 2000). The cultural dimensions in Schwartz’s (1994) framework are also clearly defined and operationalized a priori, in contrast to other frameworks that, ex post, infer cultural dimensions from correlations among diverse items and exploratory analyses (e.g. Inglehart & Baker, 2000).

Third, Schwartz’s values theory explicitly addresses the distinction between the individual and national-cultural levels of analysis. Scholars have recently challenged the notion of culture as a set of meanings and principles shared by most members of a certain society (Fischer & Schwartz, 2011). In contrast with other cultural theories, Schwartz’s conception of cultural values as a normative system that is external to individuals (but underlies the functioning of societal institutions) does not assume a high level of within-country consensus (Fischer & Schwartz, 2011; Schwartz, 2009, 2011).

All the reasons above suggest that Schwartz’s framework fits well with the topic of customer empowerment. Its bipolar dimensions capture opposing choices to three critical needs that confront most societies (Burgess & Steenkamp, 2006; Schwartz, 2006). The first dimension relates to the need to organize the relations between the individual and the group. High-autonomy cultures emphasize individuality, independence and self-expression. Affective autonomy cultures encourage individuals to act according to their own preferences. Intellectual autonomy cultures encourage individuals to develop their own opinions. In contrast, high-embeddedness cultures emphasize social relationships, group identification, respect for tradition and obedience.

The second dimension represents the need to guarantee responsible behaviors that protect the social fabric. There are two opposing ways to reach this goal. Egalitarian cultures tend to instill socially responsible behavior by inducing people to see each other as moral equals and emphasizing equality and equal distribution of power. People in such societies tend to internalize cooperation and concern with others as a life-guiding principle. Hierarchical cultures rely on an unequal distribution of power and roles as a legitimate mechanism to guarantee behaviors that protect the social fabric.

The third dimension relates to the need to manage the relations of people to society and the environment. High-mastery cultures emphasize success, daring and competence. High-harmonious cultures emphasize the need to fit into the social and natural world and the importance of behaving in a way that is congruent with the social and natural environment.

We expect culture to intensify or attenuate our hypothesized relationships for the effects of customer empowerment on non-adherence in three ways. First, as customers in high intellectual autonomy cultures are more inclined to pursue their own opinions independently, as compared with customers in low intellectual autonomy cultures (Schwartz, 2006), they should be more likely to become overconfident when exposed to expert facilitation of informational empowerment or decisional empowerment. In high-embeddedness cultures, in contrast, customers are less likely to engage in actions that may disrupt traditional roles and in-group solidarity (Burgess & Steenkamp, 2006). Thus, we expect customers in societies that emphasize embeddedness to be less likely to discount the expert’s advice, in order to avoid disrupting the customer–expert relationship, as compared to customers in societies that emphasize autonomy.

Second, when compared with customers in egalitarian societies, customers in hierarchical societies should be more likely to ascribe power to the expert because of her presumed access to superior knowledge.
and information (Burgess & Steenkamp, 2006). When customers ascribe more power to an expert, they are more likely to invest additional effort to understand and recall the expert’s advice (Tost, Gino, & Larrick, 2012). We also expect customers in hierarchical societies to be less likely to become overconfident and more likely to “comply with the obligations and rules attached to their roles and status” (Burgess & Steenkamp, 2006, p. 343). Hence, we expect the detrimental effects of customer empowerment (especially of expert facilitation of informational empowerment and decisional empowerment) on non-adherence to be less pronounced in hierarchical cultures.

Finally, we expect customers in high-mastery societies – such as the U.S. – to be more likely to perceive customer empowerment as a legitimate mechanism to enable them to control their own destiny and decisions (Markus & Schwartz, 2010). Therefore, we expect the effects of customer empowerment on non-adherence to be less detrimental, or more beneficial, in high-mastery cultures, as compared to the high-harmony cultures.

4. Data and method

4.1. Institutional context

Healthcare decisions provide a highly relevant context in which to study customer adherence to expert advice (Schwartz et al., 2011; Stremersch, 2008). In this domain, expert advice may be a therapy plan prescribed or recommended by the physician to a consumer, or patient. As stated in the introduction, therapy non-adherence generates enormous costs for society and lost sales for pharmaceutical firms, triggering significant attention in the marketing literature (Stremersch & Van Dyck, 2009; Wosinska, 2005).


We also control for other domain-specific drivers of unintentional and reasoned non-adherence to therapy advice, inspired by prior literature and befiting our theory above. In particular, we control for sociodemographics (DiMatteo, 2004), consumer–physician homophily (Dellande et al., 2004), relationship quality (Palmatier, Dant, Grewal, & Evans, 2006), duration, frequency of interaction and time since last encounter (Doney & Cannon, 1997), consumer’s perceived doctor expertise (given the role of expert power in our theory), consumer health status (DiMatteo, 2004), health motivation (Moorman & Matulich, 1993), and consumer medical knowledge (World Health Organization, 2003). Fig. 2 summarizes our conceptual framework.

4.2. Data collection method

We surveyed 11,735 consumers in Belgium, Brazil, Canada, Denmark, Estonia, France, Germany, India, Italy, Japan, The Netherlands, Poland, Portugal, Singapore, Switzerland, the UK and the US. Medical scholars have established the effectiveness of self-reports of consumers on therapy adherence (Gehi, Ali, Na, & Whooley, 2007), which correlates highly with biological measures like plasma viraemia (Walsh, Mandalia, & Gazzard, 2002). Reverse causality and common method variance are two well-known concerns with cross-sectional survey research (Rindfleisch, Malter, Ganesan, & Moorman, 2008). Section 6 provides process evidence to establish directionality. Regarding common method variance, we conducted Harmon’s one-factor test (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), and the single factor hypothesis was rejected in all countries. We also relied on different response scales and anchors (e.g. ‘never’ to ‘very often’ for non-adherence, and ‘strongly disagree’ to ‘strongly agree’ for informational empowerment), which has been shown to be an effective strategy to reduce common method bias (Rindfleisch et al., 2008). Our estimated effects also show opposite signs (e.g. decisional empowerment versus relationship quality), which is also incompatible with similar response behavior across items.

To the best of our knowledge, this is the largest study of the relationship between consumer empowerment and therapy non-adherence to date. We contracted SSI (Survey Sampling International) to execute our survey on their online panels. Recruiting and rewarding procedures for SSI panels are constantly evaluated in terms of sample representativeness and respondent’s attention and motivation.

We selected this sample of countries, because: (1) it contains sufficient cross-cultural variation; (2) consumers are free to choose their physician and typically develop repeated interactions with the same physician in each sampled country; (3) survey costs per country were not greater than $10,000. We excluded respondents that were younger than 25 or that had less than three visits with their current general practitioner, in order to guarantee respondent ability to assess the interaction with her physician and therapy non-adherence.

We constructed the original survey in English, which native speakers translated to Danish, Dutch, English, Estonian, French, German, Italian, Japanese, Polish and Portuguese. Another native speaker (the back-translator) translated the survey from his native tongue back to English. The translators and back-translators were doctoral students in social
sciences, fluent in English, attending a large European and a large American university. We discussed the translated surveys with both translators and back-translators, iteratively, until we were sure that the final survey retained exactly the same meaning in all languages. The vast majority of these graduate students were familiar with survey research methods, often through their coursework, which allowed us to discuss survey items, and their meanings, in detail.

4.3. Measurement: Individual-level constructs

In Tables A1–A4 (see Appendix A) we provide our measures, their respective sources, their reliabilities, and descriptive statistics for each focal construct and for each country. To ensure the validity of our measures, we discussed, ex ante, all items in the survey with researchers in marketing and two doctoral students in medicine to guarantee that the items were understandable and showed content validity. We typically asked the colleague to define the construct in his own words before showing her or him our proposed items and then ask for their agreement with the proposed operationalization. We pretested our purified measures in Singapore (186 subjects), The Netherlands (114 subjects) and the US (102 subjects). The pattern of answers in this pretest increased our confidence on the validity of our measures. We discarded these data and rolled-out the final survey simultaneously in all countries.

In the full sample, all scales had a reliability of at least .7, with the two-item measure for consumer health motivation as only exception (ρ = .60). We used, 5-point, multi-item scales for all constructs with the following exceptions. We used a single-item for decisional empowerment, because the measurement object (treatment choice) and its associated attribute (who is in charge of treatment choice) can both be easily envisioned by respondents (Bergkvist & Rossiter, 2007). This is also consistent with Usta and Hāblūn’s (2011) measurement of involvement of self in decision construct. We also used single items for health status (see Safran et al., 1998), age, education, gender, income, socioeconomic status, gender homophily, age homophily, relationship duration, interaction frequency and time since last visit. Unless indicated otherwise (see Appendix A), we used demeaned scores for these exogenous observed constructs.

4.4. Measurement: Country-level national culture

We obtained country-specific scores of national culture for all 17 countries from Shalom Schwartz, which are based on equally weighting scores of college students of varied majors and of schoolteachers of varied topics. These scores are similar to Schwartz (1994), but differ somewhat from these original teacher and student scores, because of the addition of new samples and updated measures (see Schwartz, 2009 for more details).

Schwartz’s (1994) cultural values theory relies on the concept of “societal means” for different cultural values, which are obtained by aggregating individual value priorities. These “societal means” capture the latent cultural orientations to which all individuals are exposed and, especially in social contexts (like customer–expert interactions), to which they tend to adapt (Fischer & Schwartz, 2011). Yet, Schwartz’s conceptualization of culture as external to the individual allows for substantial variation of individual values around these “societal means” and avoids the assumption of high within-society value consensus (Schwartz, 2011).

These cultural dimensions are therefore appropriate for cross-country comparisons but not for characterizing the values of individuals, which fits our research purposes. In cross-cultural analyses, it is important to avoid the problem of ecological fallacy. Ecological fallacy occurs when researchers assume that nation-level variables directly apply to individuals (Bond, 2002). In our case, the usage of national-level cultural dimensions is appropriate because we are interested in the role of culture as a moderator of the country-level effects of customer empowerment on non-adherence.

4.5. Model specification

In our models, i indexes respondents (i = 1,...,N; N = 11,735), c indexes countries (c = 1,...,C; C = 17), p indexes response items measuring latent constructs (p = 1,...,P; P = 28), q indexes latent endogenous constructs (q = 1,...,Q; Q = 2), and r indexes latent exogenous constructs (r = 1,...,R; R = 6). We specify our measurement equations relating the latent endogenous constructs – unintentional non-adherence (UNA) and reasoned non-adherence (RNA) – to the observed responses as follows:

\[ y_{ip} = x_{ip}^C + \lambda_p^C \cdot UNA_i + \epsilon_{ip}^C, \quad \text{for } 1 \leq p \leq 4. \quad (1) \]

\[ y_{ip} = x_{ip}^C + \lambda_p^C \cdot RNA_i + \epsilon_{ip}^C, \quad \text{for } 5 \leq p \leq 9. \quad (2) \]

And for the latent exogenous constructs as follows:

\[ y_{ip} = x_{ip}^C + \lambda_p^C \cdot \xi_r + \epsilon_{ip}^C, \quad \text{for } p > 9. \quad (3) \]

Where \( \epsilon_i \) denotes an exogenous latent variable (i.e. expert facilitation of informational empowerment (EFIE), consumer-initiated informational empowerment (CIEE), relationship quality, consumer medical knowledge, health motivation and perceived doctor expertise). We used, 5-point, multi-item scales for all constructs with the following exceptions. We used a single-item for decisional empowerment, because the measurement object (treatment choice) and its associated attribute (who is in charge of treatment choice) can both be easily envisioned by respondents (Bergkvist & Rossiter, 2007). This is also consistent with Usta and Hāblūn’s (2011) measurement of involvement of self in decision construct. We also used single items for health status (see Safran et al., 1998), age, education, gender, income, socioeconomic status, gender homophily, age homophily, relationship duration, interaction frequency and time since last visit. Unless indicated otherwise (see Appendix A), we used demeaned scores for these exogenous observed constructs.

\[ x_{ip} = \mu_p^C + \varepsilon_{ip}^C \] for all p. \quad (4)

The mean and variance of the scale usage heterogeneity component in Eq. (4) (\( \varepsilon_{ip}^C \)) is country-specific (\( \varepsilon_{ip}^C \) and \( \sigma_{ip}^C \) c = 1,...,17). Note that \( \tau_{ip}^p \) in Eq. (4), captures each respondent’s baseline tendency to score high (or low) in each of the constructs we measure. For instance, baseline tendencies for non-adherence are captured by \( \tau_{ip}^c \) where 1 ≤ p ≤ 9. For model identification, we assume that the \( \tau_{ip}^p \)’s are uncorrelated with the error terms and with the latent factors, which implies that differences in the usage of response scales are not related to respondents’ scores in the constructs being measured (see Maydeu-Olivares & Coffman, 2006).

We collect the error terms in Eqs. (1–3) in a single (P × 1) random vector of residuals, \( \Psi \), which we assume to be normally distributed as N(0, \( \Psi \)). \( \Psi \) is a (P × P) diagonal covariance matrix. The error terms are orthogonal to the latent factors.

Our structural model is defined as:

\[ UNA_i = \beta_0^{UNA} \cdot EFIE_i + \beta_1^{UNA} \cdot CIEE_i + \beta_2^{UNA} \cdot DE_i + \Gamma_1 \begin{bmatrix} \xi_i^C \\ X_i \end{bmatrix} + \delta_1^i \quad (5) \]

\[ RNA_i = \beta_0^{RNA} \cdot EFIE_i + \beta_1^{RNA} \cdot CIEE_i + \beta_2^{RNA} \cdot DE_i + \Gamma_2 \begin{bmatrix} \xi_i^C \\ X_i \end{bmatrix} + \delta_2^i \quad (6) \]

Where the \( \beta_p^c \) parameters are country-specific parameters capturing the effects of customer empowerment on unintentional and reasoned empowerment on non-adherence.

\[ a \]
non-adherence $\beta_{\text{UNA}}$ is a vector where we collect all exogenous latent variables besides the customer empowerment constructs (i.e., relationship quality, consumer medical knowledge, health motivation and perceived doctor expertise), $X_i$ is a vector where we collect all remaining control variables (i.e., all observed independent variables). Consequently, $F_i$ for $q = 1,2$, contain the structural paths corresponding to the control variables, pooled across countries. We collect all exogenous latent variables in a $\mathbb{R} \times 1$ vector $\beta_i^c$, distributed according to $\mathcal{N}(0,\Phi^c)$, where $\Phi^c$ is a $\mathbb{R} \times \mathbb{R}$ full covariance matrix. We assume the residuals, $\epsilon_i$, are independent of the latent variables and distributed $\mathcal{N}(0,\psi)$, for $q = 1,2$.

4.6. Estimation

We use Bayesian estimation, which is a more flexible approach to the estimation of theory-driven structural equation models than maximum likelihood (Muthén & Asparouhov, 2012). We specify the posterior distribution of the parameters of interest across all respondents and estimate the model simultaneously across all countries. We sample the model parameters from their posterior distributions by using the Gibbs sampler (Casella & George, 1992) with data augmentation, which allows sampling the latent constructs alongside the model parameters (Tanner & Wong, 1987). Bayesian estimation also facilitates our task of assessing the moderating effects of culture in our model. In particular, at each iteration of our Gibbs sampler, we store the correlations between each of the country-specific parameters ($\rho_{\beta}$) and the culture parameter ($\gamma_c$) (see Steenkamp & Baumgartner (1998)). Following Steenkamp and Baumgartner (1998), we test the hypothesis of full metric invariance by constraining the matrix of factor loadings to be invariant across countries. The configural model has a smaller DIC ($\text{DIC}_{\text{config}} = 611.998$) than the metric invariance model ($\text{DIC}_{\text{metric}} = 613.385$), which means we do not find support for full metric invariance ($\text{DIC}$: deviance information criterion; see Spiegelhalter, Best, Carlin, & van der Linde, 2002).

Full metric invariance is very unlikely (Steenkamp & Baumgartner, 1998, p.81) and Byrne, Shavelson, and Muthén (1989) have established that partial metric invariance is sufficient for cross-cultural equivalence and meaningful cross-national comparison. In order to understand the lack of full metric invariance, we compare the factor loadings from the measurement invariance model with those of the configural model. We first stored, at each draw, the 20 factor loadings across the 17 countries in our sample obtained from the configural model. Next, we computed the 95% credible intervals for each of these 340 loadings across the posterior draws from our MCMC chain. We then examined whether the 95% credible interval for each of the country-specific loadings from the configural model contained the posterior median of the corresponding factor loading estimated by using the metric invariance model. This was the case in 243 out of the 340 loadings (i.e. 71.5% of the loadings; see Table A3 in the Appendix A for a cross-country comparison). More importantly, when comparing the structural path estimates between the metric invariance and the configural models we saw no meaningful differences. The correlation between the focal structural paths (capturing the effects of customer empowerment on non-adherence) in the configural and metric invariance models is .99 and we do not find any significant difference across paths. In other words, in all cases, the 95% credible intervals of the structural paths in the metric invariance model contained the posterior mean of the same path according to the configural model and vice versa. Overall, these results provide strong evidence that we have sufficient cross-country equivalence to make cross-national inferences.

5. Results

5.1. Non-adherence to expert advice across countries

Fig. 3 plots the mean levels of unintentional and reasoned non-adherence across countries in our sample, computed by averaging, across the MCMC draws, the measurement intercepts ($\gamma_{\beta}$). We do not restrict the measurement intercepts across countries, since the latent means are constrained to be equal, which ensures meaningful cross-national comparison. The dashed lines in Fig. 3 represent the median levels. While there is a positive relationship between unintentional and reasoned non-adherence, the relationship is not perfect ($\rho = .80$ and a linear regression of RNA on UNA has an $R^2$ of .64). Consumers in Estonia, Japan, India and Singapore exhibit considerably higher levels of non-adherence than consumers in Denmark and The Netherlands.

5.2. Customer empowerment and non-adherence to expert advice

Table 1 presents the estimated coefficients from our multi-sample structural equation model with country-specific random effects in the measurement model capturing scale usage heterogeneity. We let all chains converge by running our models for 25,000 iterations, discarding the first 10,000 for burn-in, and using the subsequent 1500 thinned draws (we used every 10th draw to reduce autocorrelation) for posterior inference. The estimates are the posterior cross-country medians obtained from the MCMC draws (we used every 10th draw to reduce autocorrelation) for posterior inference. The estimates are the posterior cross-country medians obtained from the MCMC draws (we used every 10th draw to reduce autocorrelation).

Even though we find a positive relationship between expert facilitation of informational empowerment (EFIE) and non-adherence, the relationship was neither significant for unintentional non-adherence ($\rho_{\text{EFIE-UNA}} = .04; 95\% \text{CI} = [-.01; .09]$) nor for reasoned non-adherence ($\rho_{\text{EFIE-RNA}} = .04; 95\% \text{CI} = [.02; .09]$). These initial results do not support $H_1$ and $H_2$.

In support of $H_3$, customer-initiated informational empowerment (CIE) is associated with lower levels of unintentional non-adherence ($\rho_{\text{CIE-UNA}} = -.22; 95\% \text{CI} = [-.28; -.17]$). CIE is also associated with lower reasoned non-adherence ($\rho_{\text{CIE-RNA}} = -.16; 95\% \text{CI} = [-.21; -.11]$), which suggests that the motivational benefits of discussing, during an advising interaction, solution-relevant information that customers find self-relevant are stronger than the detrimental effects of such discussion on customer confidence.

In support of hypotheses $H_4$ and $H_5$, decisional empowerment (DE) is associated with higher unintentional non-adherence ($\rho_{\text{DE-UNA}} = .04; 95\% \text{CI} = [.03; .06]$) and with higher reasoned non-adherence ($\rho_{\text{DE-RNA}} = .08; 95\% \text{CI} = [.06; .10]$).

6. A full covariance matrix allows us to control for covariation among exogenous latent constructs (Lee, 2007).
5.3. Other drivers of non-adherence to expert advice

Table 2 presents the estimates for the control variables. Our results are in line with the findings of prior literature. We discuss several interesting paths, while a more detailed note on all effects is available from the first author upon request. The results on sociodemographics are consistent with the medical literature (DiMatteo, 2004) and recent research in marketing (Neslin et al., 2009), which find no or modest effects of sociodemographics on non-adherence.

The beneficial effects of relationship quality on therapy non-adherence are consistent with the relationship marketing literature (Geyserkens, Steenkamp, & Kumar, 1998; Morgan & Hunt, 1994). Gender homophily is associated with lower levels of unintentional non-adherence, but not reasoned non-adherence. The latter effect is consistent with prior research in marketing (Dellande et al., 2004). Reasoned non-adherence decreases with interaction frequency, which is not true for unintentional non-adherence. Reasoned non-adherence also tends to increase between visits, in line with Bowman et al. (2004). We do not find such an effect for unintentional non-adherence.

### Table 1
Effects of customer empowerment on unintentional and reasoned non-adherence.

<table>
<thead>
<tr>
<th>Posterior</th>
<th>95% Credible interval</th>
<th>Posterior</th>
<th>95% Credible interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-country median</td>
<td>Cross-country std. deviation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFIE → UNA</td>
<td>0.04</td>
<td>[0.03, 0.06]</td>
<td>0.05</td>
</tr>
<tr>
<td>EFIE → RNA</td>
<td>0.04</td>
<td>[0.02, 0.09]</td>
<td>0.05</td>
</tr>
<tr>
<td>CIE → UNA</td>
<td>0.16</td>
<td>[0.15, 0.29]</td>
<td>0.15</td>
</tr>
<tr>
<td>CIE → RNA</td>
<td>0.16</td>
<td>[0.14, 0.18]</td>
<td>0.17</td>
</tr>
<tr>
<td>DE → UNA</td>
<td>0.08</td>
<td>[0.06, 0.10]</td>
<td>0.05</td>
</tr>
<tr>
<td>DE → RNA</td>
<td>0.08</td>
<td>[0.06, 0.10]</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Acronyms: EFIE = Expert Facilitation of Informational Empowerment; CIE = Customer-Initiated Informational Empowerment; DE = Decisional Empowerment; UNA = Unintentional Non-Adherence; RNA = Reasoned Non-Adherence.

Notes: We estimate a random intercept factor analysis model (Maydeu-Olivares & Coffman, 2006) capturing systematic differences in usage of response scales. At each draw in our MCMC chain, we computed the averages and the standard deviations of the posterior means of the depicted structural paths across all countries in our sample. We stored these cross-country averages (MUs) and standard deviations (SDs). The posterior cross-country medians are the medians of these averages (MUs) across the 1500 draws we used for inference (total number of draws = 25,000; burn-in = 10,000; thinning = 10). The 95% credible intervals depict the 2.5th and the 97.5th percentiles of the distribution of these averages (MUs). We set in bold the paths whose 95% credible interval do not contain zero. The posterior cross-country standard deviations are the medians of the stored standard deviations (SDs). All endogenous and exogenous latent and observed constructs in our structural model have mean zero.

### Table 2
Control variables.

<table>
<thead>
<tr>
<th>Sociodemographics</th>
<th>Posterior median</th>
<th>95% Credible interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age → UNA</td>
<td>-0.13</td>
<td>[-0.15, -0.11]</td>
</tr>
<tr>
<td>Age → RNA</td>
<td>-0.10</td>
<td>[-0.12, -0.09]</td>
</tr>
<tr>
<td>Education → UNA</td>
<td>-0.01</td>
<td>[-0.02, -0.01]</td>
</tr>
<tr>
<td>Gender (male = 1) → RNA</td>
<td>0.04</td>
<td>[0.01, 0.09]</td>
</tr>
<tr>
<td>Income → UNA</td>
<td>-0.00</td>
<td>[-0.01, 0.00]</td>
</tr>
<tr>
<td>Income → RNA</td>
<td>-0.01</td>
<td>[-0.01, 0.00]</td>
</tr>
<tr>
<td>Socioeconomic status → UNA</td>
<td>-0.01</td>
<td>[-0.02, 0.01]</td>
</tr>
<tr>
<td>Socioeconomic status → RNA</td>
<td>-0.01</td>
<td>[-0.03, 0.01]</td>
</tr>
<tr>
<td>Relationship quality → UNA</td>
<td>-0.06</td>
<td>[-0.11, -0.00]</td>
</tr>
<tr>
<td>Relationship quality → RNA</td>
<td>-0.03</td>
<td>[-0.08, -0.02]</td>
</tr>
<tr>
<td>Gender homophily → UNA</td>
<td>-0.01</td>
<td>[0.02, -0.01]</td>
</tr>
<tr>
<td>Gender homophily → RNA</td>
<td>-0.00</td>
<td>[-0.02, 0.01]</td>
</tr>
<tr>
<td>Relationship duration → UNA</td>
<td>-0.03</td>
<td>[-0.05, -0.02]</td>
</tr>
<tr>
<td>Relationship duration → RNA</td>
<td>-0.02</td>
<td>[-0.04, -0.01]</td>
</tr>
<tr>
<td>Interaction frequency → UNA</td>
<td>-0.01</td>
<td>[-0.02, -0.01]</td>
</tr>
<tr>
<td>Interaction frequency → RNA</td>
<td>-0.04</td>
<td>[-0.05, -0.02]</td>
</tr>
<tr>
<td>Time since last visit → UNA</td>
<td>0.00</td>
<td>[0.01, -0.02]</td>
</tr>
<tr>
<td>Time since last visit → RNA</td>
<td>0.02</td>
<td>[0.03, -0.01]</td>
</tr>
</tbody>
</table>

#### Health drivers
| Consumer medical knowledge → UNA | -0.21 | [-0.24, -0.18] |
| Consumer medical knowledge → RNA | -0.17 | [-0.20, -0.15] |
| Health status → UNA | -0.07 | [-0.09, -0.05] |
| Health status → RNA | -0.02 | [-0.04, -0.00] |
| Health motivation → UNA | -0.39 | [-0.43, -0.34] |
| Health motivation → RNA | -0.27 | [-0.31, -0.23] |
| Doctor expertise → UNA | -0.17 | [-0.25, -0.08] |
| Doctor expertise → RNA | -0.26 | [-0.35, -0.18] |

Acronyms: EFIE = Expert Facilitation of Informational Empowerment; CIE = Customer-Initiated Informational Empowerment; DE = Decisional Empowerment; UNA = Unintentional Non-Adherence; RNA = Reasoned Non-Adherence.

Note: For model stability and identification, the structural paths for control variables were estimated pooled across countries.

6. Process evidence

We now discuss process evidence on the effects of customer empowerment on unintentional and reasoned non-adherence. We use customer-centered communication quality (i.e., the extent to which the customer believes that her doctor spends sufficient time, during an
advising interaction, sharing clear and understandable information with her, see Kao, Green, Zaslavsky, Koplan, & Cleary, (1998) and locus of control (the customer’s confidence in her own ability to cure herself, see Moorman & Matulich, 1993) as mediators (see Table 3). We now discuss the influence of these mediators on non-adherence, after which we turn to the influence of empowerment on these mediators.

In line with our expectations, high customer-centered communication quality is associated with lower levels of unintentional non-adherence (γ_{cultural:UNA} = −.26; 95% CI = [−.31; −.22]) and reasoned non-adherence (γ_{cultural:RNA} = −.29; 95% CI = [−.34; −.25]). Also as theorized, high locus of control – i.e. our proxy for customer overconfidence is associated with higher levels of unintentional (γ_{cultural:UNA} = .03; 95% CI = [.02; .05]) and reasoned non-adherence (γ_{cultural:RNA} = .09; 95% CI = [.07; .11]).

Expert facilitation of informational empowerment (EFIE) is associated with worse customer-centered communication quality (β_{EFIE-CQUAL} = .67; 95% CI = [.64, .70]), but this effect is offset by a direct effect on unintentional non-adherence (β_{EFIE-UNA} = .25; 95% CI = [1.18, .31]). This is consistent with the logic under H1: That is, unrequested solution-relevant information makes the advice harder to recall and may crowd out other pieces of information that may be more relevant to stimulate adherence (Epstein et al., 2010), offsetting the beneficial impact of EFIE on customer-centered communication quality. EFIE is negatively, but insignificantly, associated to locus of control (β_{EFIE-LOCUS} = −.05; 95% CI = [−.12; .02]) and positively associated with reasoned non-adherence (β_{EFIE-RNA} = .25; 95% CI = [20; .31]). This suggests that EFIE increases customers’ tendency to egocentrically discount the expert’s advice, in line with the behavioral mechanism underlying H2.

Customer-initiated informational empowerment (CIIE) is associated with worse customer-centered communication quality (β_{CIIE-CQUAL} = −.06; 95% CI = [−.10; −.03]) but also with lower unintentional non-adherence (β_{CIIE-UNA} = −.23; 95% CI = [−.29; −.18]). These two effects are consistent with the relations we theorized in H2 and Schwartz’s country-specific cultural dimensions. This analysis revealed that the effects of decisional empowerment and, to a lesser extent, of customer-initiated informational empowerment and expert facilitation of informational empowerment on non-adherence are moderated by culture. Culture is a stronger moderator of RNA (ten posterior correlations with 95% credible intervals not containing zero) than of UNA (two posterior correlations with 95% credible intervals not containing zero). We now discuss the moderating effects of culture on the relationship between decisional empowerment and non-adherence.

In high-embeddedness cultures (ρ_{CIE > RNA},EMBEDDEDNESS = −.31, 95% CI = [−.57; −.00]); high-hierarchy cultures (ρ_{CIE > RNA},HIERARCHY = −.34, 95% CI = [−.60; −.01]), and high-mastery cultures (ρ_{CIE > RNA},MASTERY = −.40, 95% CI = [−.67; −.06]), decisional empowerment is less detrimental as it increases reasoned non-adherence less than in other cultures. In high-intellectual autonomy cultures (ρ_{CIE > RNA},INTA = .44, 95% CI = [.15; .69]) and harmonious cultures (ρ_{CIE > RNA},HARMONY = .57, 95% CI = [.32; .77]), decisional

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effects of the mediators on non-adherence</td>
</tr>
<tr>
<td>Communication quality → UNA</td>
</tr>
<tr>
<td>Communication quality → RNA</td>
</tr>
<tr>
<td>Health locus of control → UNA</td>
</tr>
<tr>
<td>Health locus of control → RNA</td>
</tr>
</tbody>
</table>

**Effects of empowerment on the mediators**

| EFIE → Communication quality | .67 | [.64, .70] | .15 |
| EFIE → Health locus of control | −.05 | [−1.20, .02] | .22 |
| CIIE → Communication quality | −.06 | [−.10; −.03] | .19 |
| CIIE → Health locus of control | .24 | [.15, .31] | .28 |
| DE → Communication quality | −.03 | [−.04, −.02] | .03 |
| DE → Health locus of control | .11 | [.04, .16] | .14 |

**Direct effects**

| EFIE → UNA | .25 | [.18, .31] | .16 |
| EFIE → RNA | .25 | [.20, .31] | .15 |
| CIIE → UNA | −.23 | [−.29, −.18] | .16 |
| CIIE → RNA | −.16 | [−.22, −.11] | .17 |
| DE → UNA | .04 | [.02, .05] | .05 |
| DE → RNA | .07 | [.05, .08] | .06 |

Acronyms: EFIE = Expert Facilitation of Informational Empowerment; CIIE = Customer-Initiated Informational Empowerment; DE = Decisional Empowerment; UNA = Unintentional Non-Adherence; RNA = Reasoned Non-Adherence. Note: The model includes the same set of control variables used in our main model. The full set of parameter estimates is available upon request.
empowerment increases reasoned non-adherence more than in other cultures. We also find that in harmonious cultures, decisional empowerment increases unintentional non-adherence more than in other cultures (EFIE → UNA-HARMONY = .36, 95% CI = [.05; .62]). These effects are in line with our theory-driven expectations. For instance, when compared with customers in high-mastery societies, customers in harmonious societies should perceive decisional empowerment as more incongruent with the expected roles of customers and experts. Higher perceived incongruence, in turn, may magnify the detrimental effects of decisional empowerment on unintentional and reasoned non-adherence.

The effect of expert facilitation of informational empowerment (EFIE) on reasoned non-adherence is less detrimental in high-hierarchy countries (EFIE > RNA, HIERARCHY = −.56, 95% CI = [−.79; −.16]) but more detrimental in high-egalitarianism cultures (EFIE > RNA, Egalitarianism = .40, 95% CI = [.02; .63]). The effect of EFIE on unintentional non-adherence is also less detrimental in high-hierarchy countries (EFIE > UNA-HARMONY = −.45, 95% CI = [−.71; −.08]), as compared to customers in less hierarchical countries. In high-hierarchy cultures, customers are less likely to engage in behaviors that threaten the expert’s role and, thus, EFIE has less detrimental effects than in other countries (e.g. high-egalitarianism).

Finally, the beneficial effect of customer-initiated informational empowerment (CIE) on reasoned non-adherence is weaker in high-affective autonomy and high-intellectual autonomy cultures than in other cultures (CIE > RNA, AFF-AUTONOMY = .40, 95% CI = [.13; .64]; CIE > RNA, INT.AUTONOMY = .35, 95% CI = [.03; .59]) but stronger in high-embeddedness cultures (CIE > RNA, EMBEDDEDNESS = −.43, 95% CI = [−.66; −.11]). Compared with customers in high-embeddedness cultures, customers in more autonomous cultures may have a tendency to be vocal (high CIE) but also to follow their own opinion even if that entails discounting an expert’s opinion (high RNA).

For example, in high-mastery cultures, such as the U.S., decisional empowerment triggers less customer overconfidence and thus less reasoned non-adherence. Exploring such cross-cultural heterogeneity allowed us to better understand in which cultures empowerment may have the largest or smallest impact on non-adherence.

### 8.1. Implications

These findings provide important and counterintuitive insights. The current thinking among many scholars is that shared informed autonomy (high decisional and informational empowerment) is the customer–expert decision-making model that minimizes non-adherence to expert advice (Epstein et al., 2004; Macfarlane, 2008). Financial and tax advisors, lawyers, doctors and management consultants – to name just a few – routinely consider whether accommodating the whims and opinions of their customers (versus maintaining a strong opinion and decision control) would help them achieve better results (Usta & Häubl, 2011), in particular higher customer adherence to their advice (Epstein et al., 2004; Quill & Brody, 1996).

In contrast with this view, we find that customer-driven informed delegation is the model that minimizes non-adherence to expert advice. The underpinnings of this model are that: (1) decision power should remain with the expert if the expert wishes the customer to adhere to his advice, (2) customers should be allowed to ask questions and offer their opinion, and (3) experts should not proactively facilitate informational empowerment.

In the specific case of doctor–patient interactions – the institutional context of our empirical analysis – these findings are particularly timely. From the famous paternalistic scenes in the movie “Patch Adams,” the medical decision-making model is now undergoing increasing pressure to be more consumer-centric. In light of our findings, the concern that consumer-centricity may in practice turn to healthcare consumerism and reduce healthcare quality (Camacho, 2014; Starkey, 2003), seems valid for treatment non-adherence. In the optimal model – customer-driven informed delegation – the physician acts as an agent to whom customers delegate authority (a feature also present in the paternalistic model) and is responsive, but not proactive, to exchange solution-relevant information (a feature that is not present in the paternalistic model).

Cross-national heterogeneity in the magnitude of our effects allows us to offer some culturally-specific implications. In particular, sharing more decision power with customers would be less detrimental for experts in the US (a culture that emphasizes mastery and self-assertion) than for experts in many Western European countries such as Denmark, France, Germany, and the Netherlands (cultures that emphasize harmony and intellectual autonomy).

### 8.2. Future research

Our study has several limitations that can open new avenues for future research. First, future research using revealed customer adherence data, for instance, from script refills, holds great promise, because it shows greater external validity. On the other hand, such data may...
contain less detail (e.g. no distinction between reasoned and unintentional non-adherence), possibly contain self-selection mechanisms (e.g. most patient-monitoring programs are opt-in) and be hard to obtain.

Second, in this paper, we study customer non-adherence as a behavioral trait (in line for instance with Bowman et al., 2004; DiMatteo et al., 1993). Despite this tradition, it would be interesting if future research would look into context-specific motivations for adherence. Most studies have also focused on patient adherence to physician advice. It would be interesting if future work in marketing explored customer non-adherence in contexts beyond healthcare, such as consulting, financial or tax advice, and legal advice.

Third, in our cultural analyses, we rely on country-level cultural scores and test whether these scores predict variation in country-specific effects of customer empowerment on adherence to expert advice. Future research could rely on individual-level value scores to explore within-country value heterogeneity and test the sensitivity of our results to the unit of analysis chosen for cultural inferences.

Fourth, future research could also explore behavioral interventions aimed at reducing unintentional and/or reasoned non-adherence to treatment advice. For instance, Adhere.IO is a behavioral diagnostic invented at MIT that uses lateral flow technology – the technology used in pregnancy tests – to verify, remotely, if a patient took her drugs on time and reward those who accurately follow the therapy advice (Gomez-Marquez, 2013). Future studies could help optimize this type of behavioral interventions to maximize reduction of unintentional and/or reasoned non-adherence.

Fifth, in this study we assumed that the customer seeks the advice of a single expert. Research on advice-taking, however, suggests that integrating the advice of multiple experts may improve customers’ decisions (Bonaccio & Dalal, 2006; Budescu, Rantilla, Yu, & Karelitz, 2003). Future research could thus examine how customers integrate and weigh the advice from multiple experts possibly with distinct decision-making styles.

Sixth, there are also many situations where adherence to expert advice is not an individual, but a group decision. For instance, when lawyers or management consultants advise an executive committee on a litigation or business strategy, adherence to the expert’s advice is determined through negotiation among the members of the executive committee. Future research may explore advice-giving to multiple agents in the same decision-making unit and the optimality of different customer–expert decision-making models in such contexts.

Seventh, existing research on dual-process models has identified several antecedents of people’s tendency to engage in heuristic or systematic information processing (e.g. Chaiken et al., 1989). Future research could further explore customer, expert and customer–expert interaction characteristics that may trigger the activation of these different types of information processing modes and influence customer adherence to expert advice.

In general, the present paper may inform policy discussions on patient empowerment. It may also guide experts on how to engage with their customers, to the extent that they want their customers to adhere to their advice. Finally, it also may be informative for customers, because they may themselves suffer from not adhering to expert advice.

Appendix A. Measures and metric invariance

Table A1
Constructs and measures [source]

<table>
<thead>
<tr>
<th>Constructs and measures [source]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unintentional non-adherence (α = .84) [DiMatteo et al. (1993)]: Please tell us how often you can imagine yourself …</td>
</tr>
<tr>
<td>1. …forgetting to take your medicines?</td>
</tr>
<tr>
<td>2. …having a hard time doing what your doctor suggested you to do?</td>
</tr>
<tr>
<td>3. …being unable to do what was necessary to follow your doctor’s treatment plans?</td>
</tr>
<tr>
<td>4. …missing taking your medications because you were away from home or busy with other things?</td>
</tr>
<tr>
<td>Reasoned non-adherence (α = .87) [DiMatteo et al. (1993)]: Please tell us how often you can imagine yourself missing taking your medications because…</td>
</tr>
<tr>
<td>1. … you seemed to need less medicine?</td>
</tr>
<tr>
<td>2. … you didn’t believe in the treatment your doctor was recommending you?</td>
</tr>
<tr>
<td>3. … you wanted to avoid side effects or felt like the drug was toxic or harmful?</td>
</tr>
<tr>
<td>4. … you wanted to try alternative therapies (e.g. herbalist, homeopathic or acupuncture treatments…)?</td>
</tr>
<tr>
<td>5. …the medication was too expensive?</td>
</tr>
<tr>
<td>Response scale for non-adherence: 1 = “never,” 2 = “rarely,” 3 = “sometimes,” 4 = “often,” 5 = “very often”</td>
</tr>
<tr>
<td>Expert facilitation of informational empowerment (α = .83) [Kao et al. (1998); Lerman et al. (1990)]: Please read each of the statements below and indicate to what extent it describes your own experience with your doctor.</td>
</tr>
<tr>
<td>1. My doctor asks me about how my family or living situation might affect my health.</td>
</tr>
<tr>
<td>2. My doctor shares with me the risks and benefits associated with alternative treatment options.</td>
</tr>
<tr>
<td>3. My doctor asks me what I believe is causing my medical symptoms.</td>
</tr>
<tr>
<td>4. My doctor encourages me to give my opinion about medical treatments.</td>
</tr>
<tr>
<td>Customer-initiated informational empowerment (α = .74) [Lerman et al. (1990)]: Please read each of the statements below and indicate to what extent it describes your own experience with your doctor.</td>
</tr>
<tr>
<td>1. I ask my doctor to explain to me the treatments or procedures in detail.</td>
</tr>
<tr>
<td>2. I ask my doctor a lot of questions about my medical symptoms.</td>
</tr>
<tr>
<td>3. I give my opinion (agreement or disagreement) about the types of test or treatment that my doctor orders.</td>
</tr>
<tr>
<td>Decisional empowerment [Similar to Usta and Häubl (2011)]: Who possesses more power in treatment decisions, that is, who has more influence in determining the treatment(s) you follow?</td>
</tr>
<tr>
<td>Response scale: 1 = “my doctor has more power,” 2 = “my doctor has slightly more power,” 3 = “my doctor and I have about the same ability,” 4 = “I have slightly more power,” 5 = “I have more power.”</td>
</tr>
<tr>
<td>Communication quality (α = .89) [Kao et al. (1998)]: Please read each of the statements below and indicate to what extent it describes your own experience with your doctor.</td>
</tr>
<tr>
<td>1. When I ask questions to my doctor, I get answers that are understandable.</td>
</tr>
<tr>
<td>2. My doctor gives me enough time to explain the reasons for my visit.</td>
</tr>
<tr>
<td>3. My doctor takes enough time to answer my questions.</td>
</tr>
<tr>
<td>Health locus of control [Item from Moorman and Matulich (1993)]:</td>
</tr>
<tr>
<td>1. I have a lot of confidence in my ability to cure myself once I get sick.</td>
</tr>
</tbody>
</table>

Response scale: 1 = “strongly disagree,” 2 = “disagree,” 3 = “neutral,” 4 = “agree,” 5 = “strongly agree”
For two-item scales we report Pearson’s correlation coefficients between them. Acronyms: UNA = Unintentional Non-Adherence; RNA = Reasoned Non-Adherence. EFIE = Expert Facilitation of Informational Empowerment; CIIE = Customized Informational Empowerment. Table A1 reports reliabilities per country:

<table>
<thead>
<tr>
<th>Country</th>
<th>N</th>
<th>UNA</th>
<th>RNA</th>
<th>EFIE</th>
<th>CIIE</th>
<th>QC</th>
<th>RQ</th>
<th>CMK*</th>
<th>HM**</th>
<th>PDE**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>669</td>
<td>.813</td>
<td>.817</td>
<td>.803</td>
<td>.703</td>
<td>.898</td>
<td>.782</td>
<td>.774</td>
<td>.628</td>
<td>.807</td>
</tr>
<tr>
<td>Brazil</td>
<td>785</td>
<td>.864</td>
<td>.864</td>
<td>.788</td>
<td>.674</td>
<td>.858</td>
<td>.791</td>
<td>.840</td>
<td>.501</td>
<td>.793</td>
</tr>
<tr>
<td>Canada</td>
<td>540</td>
<td>.830</td>
<td>.871</td>
<td>.841</td>
<td>.721</td>
<td>.896</td>
<td>.881</td>
<td>.788</td>
<td>.550</td>
<td>.843</td>
</tr>
<tr>
<td>Denmark</td>
<td>570</td>
<td>.777</td>
<td>.828</td>
<td>.830</td>
<td>.754</td>
<td>.916</td>
<td>.829</td>
<td>.752</td>
<td>.624</td>
<td>.777</td>
</tr>
<tr>
<td>Estonia</td>
<td>523</td>
<td>.792</td>
<td>.778</td>
<td>.842</td>
<td>.660</td>
<td>.839</td>
<td>.807</td>
<td>.779</td>
<td>.575</td>
<td>.765</td>
</tr>
<tr>
<td>France</td>
<td>776</td>
<td>.775</td>
<td>.820</td>
<td>.823</td>
<td>.711</td>
<td>.875</td>
<td>.818</td>
<td>.672</td>
<td>.442</td>
<td>.842</td>
</tr>
<tr>
<td>Germany</td>
<td>783</td>
<td>.824</td>
<td>.872</td>
<td>.888</td>
<td>.753</td>
<td>.926</td>
<td>.822</td>
<td>.911</td>
<td>.595</td>
<td>.809</td>
</tr>
<tr>
<td>India</td>
<td>521</td>
<td>.853</td>
<td>.909</td>
<td>.824</td>
<td>.706</td>
<td>.817</td>
<td>.850</td>
<td>.640</td>
<td>.547</td>
<td>.810</td>
</tr>
<tr>
<td>Italy</td>
<td>818</td>
<td>.862</td>
<td>.853</td>
<td>.842</td>
<td>.724</td>
<td>.897</td>
<td>.876</td>
<td>.722</td>
<td>.528</td>
<td>.830</td>
</tr>
<tr>
<td>Japan</td>
<td>758</td>
<td>.846</td>
<td>.900</td>
<td>.770</td>
<td>.819</td>
<td>.830</td>
<td>.828</td>
<td>.645</td>
<td>.659</td>
<td>.691</td>
</tr>
<tr>
<td>Poland</td>
<td>760</td>
<td>.873</td>
<td>.873</td>
<td>.840</td>
<td>.721</td>
<td>.916</td>
<td>.775</td>
<td>.830</td>
<td>.755</td>
<td>.816</td>
</tr>
<tr>
<td>Portugal</td>
<td>524</td>
<td>.877</td>
<td>.863</td>
<td>.856</td>
<td>.716</td>
<td>.892</td>
<td>.864</td>
<td>.784</td>
<td>.524</td>
<td>.829</td>
</tr>
<tr>
<td>Singapore</td>
<td>815</td>
<td>.895</td>
<td>.880</td>
<td>.803</td>
<td>.701</td>
<td>.898</td>
<td>.865</td>
<td>.761</td>
<td>.677</td>
<td>.844</td>
</tr>
<tr>
<td>Switzerland</td>
<td>547</td>
<td>.785</td>
<td>.837</td>
<td>.821</td>
<td>.716</td>
<td>.877</td>
<td>.786</td>
<td>.757</td>
<td>.549</td>
<td>.796</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>795</td>
<td>.766</td>
<td>.820</td>
<td>.812</td>
<td>.770</td>
<td>.900</td>
<td>.851</td>
<td>.717</td>
<td>.625</td>
<td>.826</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>781</td>
<td>.840</td>
<td>.870</td>
<td>.841</td>
<td>.726</td>
<td>.903</td>
<td>.879</td>
<td>.785</td>
<td>.532</td>
<td>.845</td>
</tr>
<tr>
<td>United States</td>
<td>770</td>
<td>.824</td>
<td>.889</td>
<td>.816</td>
<td>.728</td>
<td>.900</td>
<td>.869</td>
<td>.784</td>
<td>.660</td>
<td>.828</td>
</tr>
<tr>
<td>Pooled</td>
<td>11,735</td>
<td>.844</td>
<td>.870</td>
<td>.833</td>
<td>.736</td>
<td>.802</td>
<td>.833</td>
<td>.765</td>
<td>.595</td>
<td>.822</td>
</tr>
</tbody>
</table>

Nr Items: 4 5 4 3 3 6 2 2

Table A2

Scale reliabilities per country.

<table>
<thead>
<tr>
<th>Country</th>
<th>N</th>
<th>UNA*</th>
<th>RNA*</th>
<th>EFE*</th>
<th>CIIE*</th>
<th>QC</th>
<th>RQ</th>
<th>CMK*</th>
<th>HM**</th>
<th>PDE*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>669</td>
<td>.815</td>
<td>.817</td>
<td>.803</td>
<td>.703</td>
<td>.898</td>
<td>.782</td>
<td>.774</td>
<td>.628</td>
<td>.807</td>
</tr>
<tr>
<td>Brazil</td>
<td>785</td>
<td>.864</td>
<td>.864</td>
<td>.788</td>
<td>.674</td>
<td>.858</td>
<td>.791</td>
<td>.840</td>
<td>.501</td>
<td>.793</td>
</tr>
<tr>
<td>Canada</td>
<td>540</td>
<td>.830</td>
<td>.871</td>
<td>.841</td>
<td>.721</td>
<td>.896</td>
<td>.881</td>
<td>.788</td>
<td>.550</td>
<td>.843</td>
</tr>
<tr>
<td>Denmark</td>
<td>570</td>
<td>.777</td>
<td>.828</td>
<td>.830</td>
<td>.754</td>
<td>.916</td>
<td>.829</td>
<td>.752</td>
<td>.624</td>
<td>.777</td>
</tr>
<tr>
<td>Estonia</td>
<td>523</td>
<td>.792</td>
<td>.778</td>
<td>.842</td>
<td>.660</td>
<td>.839</td>
<td>.807</td>
<td>.779</td>
<td>.575</td>
<td>.765</td>
</tr>
<tr>
<td>France</td>
<td>776</td>
<td>.775</td>
<td>.820</td>
<td>.823</td>
<td>.711</td>
<td>.875</td>
<td>.818</td>
<td>.672</td>
<td>.442</td>
<td>.842</td>
</tr>
<tr>
<td>Germany</td>
<td>783</td>
<td>.824</td>
<td>.872</td>
<td>.888</td>
<td>.753</td>
<td>.926</td>
<td>.822</td>
<td>.911</td>
<td>.595</td>
<td>.809</td>
</tr>
<tr>
<td>India</td>
<td>521</td>
<td>.853</td>
<td>.909</td>
<td>.824</td>
<td>.706</td>
<td>.817</td>
<td>.850</td>
<td>.640</td>
<td>.547</td>
<td>.810</td>
</tr>
<tr>
<td>Italy</td>
<td>818</td>
<td>.862</td>
<td>.853</td>
<td>.842</td>
<td>.724</td>
<td>.897</td>
<td>.876</td>
<td>.722</td>
<td>.528</td>
<td>.830</td>
</tr>
<tr>
<td>Japan</td>
<td>758</td>
<td>.846</td>
<td>.900</td>
<td>.770</td>
<td>.819</td>
<td>.830</td>
<td>.828</td>
<td>.645</td>
<td>.659</td>
<td>.691</td>
</tr>
<tr>
<td>Poland</td>
<td>760</td>
<td>.873</td>
<td>.873</td>
<td>.840</td>
<td>.721</td>
<td>.916</td>
<td>.775</td>
<td>.830</td>
<td>.755</td>
<td>.816</td>
</tr>
<tr>
<td>Portugal</td>
<td>524</td>
<td>.877</td>
<td>.863</td>
<td>.856</td>
<td>.716</td>
<td>.892</td>
<td>.864</td>
<td>.784</td>
<td>.524</td>
<td>.829</td>
</tr>
<tr>
<td>Singapore</td>
<td>815</td>
<td>.895</td>
<td>.880</td>
<td>.803</td>
<td>.701</td>
<td>.898</td>
<td>.865</td>
<td>.761</td>
<td>.677</td>
<td>.844</td>
</tr>
<tr>
<td>Switzerland</td>
<td>547</td>
<td>.785</td>
<td>.837</td>
<td>.821</td>
<td>.716</td>
<td>.877</td>
<td>.786</td>
<td>.757</td>
<td>.549</td>
<td>.796</td>
</tr>
</tbody>
</table>

Notes:

- For multi-item scales with more than two items we report Cronbach’s alpha as our measure of scale reliability.
- For two-item scales we report Pearson’s correlation coefficient as our measure of scale reliability.
Table A3
Cross-country comparison of factor loadings between metric invariance and configural models.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Belgium</th>
<th>Brazil</th>
<th>Canada</th>
<th>Denmark</th>
<th>Estonia</th>
<th>France</th>
<th>Germany</th>
<th>India</th>
<th>Italy</th>
<th>Japan</th>
<th>Poland</th>
<th>Portugal</th>
<th>Singapore</th>
<th>Switzerland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctor expertise</td>
<td>4.37</td>
<td>4.53</td>
<td>4.68</td>
<td>4.76</td>
<td>4.51</td>
<td>4.68</td>
<td>4.81</td>
<td>4.38</td>
<td>4.78</td>
<td>4.34</td>
<td>4.37</td>
<td>4.35</td>
<td>4.68</td>
<td>4.68</td>
</tr>
<tr>
<td>Health motivation</td>
<td>3.70</td>
<td>3.67</td>
<td>3.87</td>
<td>3.79</td>
<td>3.59</td>
<td>3.46</td>
<td>3.79</td>
<td>3.48</td>
<td>3.57</td>
<td>3.61</td>
<td>3.38</td>
<td>3.48</td>
<td>3.92</td>
<td>3.67</td>
</tr>
<tr>
<td>Relationship quality</td>
<td>2.01</td>
<td>1.96</td>
<td>2.01</td>
<td>2.07</td>
<td>2.07</td>
<td>2.02</td>
<td>2.01</td>
<td>1.91</td>
<td>1.97</td>
<td>2.01</td>
<td>1.99</td>
<td>1.97</td>
<td>2.02</td>
<td>1.94</td>
</tr>
<tr>
<td>Relationship quality</td>
<td>2.01</td>
<td>1.96</td>
<td>2.01</td>
<td>2.07</td>
<td>2.07</td>
<td>2.02</td>
<td>2.01</td>
<td>1.91</td>
<td>1.97</td>
<td>2.01</td>
<td>1.99</td>
<td>1.97</td>
<td>2.02</td>
<td>1.94</td>
</tr>
</tbody>
</table>

Table A4 (continued)
Country-specific descriptive statistics.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Belgium</th>
<th>Brazil</th>
<th>Canada</th>
<th>Denmark</th>
<th>Estonia</th>
<th>France</th>
<th>Germany</th>
<th>India</th>
<th>Italy</th>
<th>Japan</th>
<th>Poland</th>
<th>Portugal</th>
<th>Singapore</th>
<th>Switzerland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctor expertise</td>
<td>4.37</td>
<td>4.53</td>
<td>4.68</td>
<td>4.76</td>
<td>4.51</td>
<td>4.68</td>
<td>4.81</td>
<td>4.38</td>
<td>4.78</td>
<td>4.34</td>
<td>4.37</td>
<td>4.35</td>
<td>4.68</td>
<td>4.68</td>
</tr>
<tr>
<td>Health motivation</td>
<td>3.70</td>
<td>3.67</td>
<td>3.87</td>
<td>3.79</td>
<td>3.59</td>
<td>3.46</td>
<td>3.79</td>
<td>3.48</td>
<td>3.57</td>
<td>3.61</td>
<td>3.38</td>
<td>3.48</td>
<td>3.92</td>
<td>3.67</td>
</tr>
<tr>
<td>Relationship quality</td>
<td>2.01</td>
<td>1.96</td>
<td>2.01</td>
<td>2.07</td>
<td>2.07</td>
<td>2.02</td>
<td>2.01</td>
<td>1.91</td>
<td>1.97</td>
<td>2.01</td>
<td>1.99</td>
<td>1.97</td>
<td>2.02</td>
<td>1.94</td>
</tr>
<tr>
<td>Relationship quality</td>
<td>2.01</td>
<td>1.96</td>
<td>2.01</td>
<td>2.07</td>
<td>2.07</td>
<td>2.02</td>
<td>2.01</td>
<td>1.91</td>
<td>1.97</td>
<td>2.01</td>
<td>1.99</td>
<td>1.97</td>
<td>2.02</td>
<td>1.94</td>
</tr>
</tbody>
</table>

Acronyms: UNA = Unintentional Non-Adherence; RNA = Reasoned Non-Adherence; EFIE = Expert Facilitation of Informational Empowerment; CIE = Customer-Initiated Informational Empowerment; DE = Decisional Empowerment.

Appendix B. Supplementary data
Estimation code for this article can be found online at http://www.runmycode.org. Interested scholars may contact either the corresponding author or IRJM’s editorial office in order to request the dataset.

References


Specifying formatively-measured constructs in endogenous positions in structural equation models: Caveats and guidelines for researchers

Dirk Temme a,⁎, Adamantios Diamantopoulos b,1, Vanessa Pfegfeidel a,2

a Methods in Empirical Economic and Social Research, University of Wuppertal, D-42097 Wuppertal, Germany
b International Marketing, University of Vienna, A-1210 Vienna, Austria

Article info

Article history:
First received in July 21, 2012 and was under review for 7 months
Available online 24 April 2014

Keywords:
Formatively-measured constructs
Endogenous formative indicators
Covariance structure analysis
Partial least squares

Abstract

Formatively-measured constructs (FMCs) are increasingly used in marketing research as well as in other disciplines. Although constructs operationalized by means of formative indicators have mostly been placed in exogenous positions in structural equation models, they also frequently occupy structurally endogenous positions. The vast majority of studies specifying endogenously-positioned FMCs have followed the common practice of modeling the impact of antecedent (predictor) constructs directly on the focal FMC without specifying indirect effects via the formative indicators. However, while widespread even in top journals, this practice is highly problematic as it can lead to biased parameter estimates, erroneous total effects, and questionable conclusions. As a result both theory development and empirically-based managerial recommendations are likely to suffer. Against this background, the authors offer appropriate modeling guidelines to ensure that a conceptually sound and statistically correct model specification is obtained when a FMC occupies an endogenous position. The proposed guidelines are illustrated using both covariance structure analysis (CSA) and partial least squares (PLS) methods and are applied to a real-life empirical example. Implications for researchers are considered and ‘good practice’ recommendations offered.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

In recent years, formative measurement models whereby “the direction of causality flows from the indicators to the latent construct, and the indicators, as a group, jointly determine the conceptual and empirical meaning of the construct” (Jarvis, MacKenzie, & Podsakoff, 2003, p. 201) have increasingly been used in marketing studies to operationalize constructs as diverse as e-service quality (Collier & Bienstock, 2006a), relationship value (Ulaga & Eggert, 2006), retailer equity (Arnett, Laverie, & Meiers, 2003) and strategic responsiveness (Nakata, Zhu, & Izberk-Bilgin, 2011) to name but a few.

While there exists an impressive body of literature on formative measurement models dealing with such issues as model specification and identification (for a review, see Diamantopoulos, Riefler, & Roth, 2008), the focus has, with few exceptions, been on structurally exogenous formatively-measured constructs (FMCs). By ‘structurally exogenous’ we mean that the focal construct is only influenced by its assigned formative indicators but has no further observed causes. FMCs, however, might also occupy an endogenous position in the structural equation model representing a researcher’s theory. Here, in addition to the formative indicators, one or more antecedent constructs are also hypothesized to impact the FMC.

Unfortunately, FMCs cannot be placed in structurally endogenous positions in the same way as their reflectively-measured counterparts, that is, by simply specifying a direct link from the antecedent to the endogenous construct. As will be subsequently demonstrated, doing so and failing to realize that the antecedent construct’s impact on the endogenous FMC should in fact be captured by its indirect effect via the formative indicators, will almost always lead to biased structural parameters, incorrect effect sizes and, ultimately, erroneous study conclusions. As a result, theory development and the generation of empirically-based managerial insights are both likely to suffer.

Although, in recent years, a few authors (Cadogan & Lee, 2013; Temme & Hildebrandt, 2006; Wiley, 2005) have warned against linking FMCs and their antecedents solely by direct construct-level paths (i.e., without further links to the formative indicators), such warnings have remained unheeded as a literature review on the use of endogenous FMCs in top journals demonstrates (see Section 2). The prevalence of misspecification in the literature can partly be attributed to the fact that the (erroneous) practice of modeling the influence of antecedent constructs on endogenous FMCs only by a direct link at the construct level can also be found in methodological papers (e.g., Jarvis et al.,...
2003; MacCallum & Browne, 1993; MacKenzie, Podsakoff, & Jarvis, 2005) and partly to the absence of concrete guidelines on how to correctly model such relationships.

The purpose of the current paper is to offer such guidelines so as to enable marketing researchers to correctly assess the influence of explanatory variables on FMCs under both covariance structure analysis (CSA) and partial least squares (PLS) perspectives. In doing so, we highlight several conceptual and methodological issues and also present an empirical illustration of ‘good practice’.

2. Misspecification of endogenously-positioned FMCs

Fig. 1 shows how endogenous FMCs are typically modeled in empirical literature. Here EMPLOR is the FMC with formative indicators x1 and x2, while x1 is the antecedent construct with reflective indicators x1–x3. The relationship between x1 and y1 is captured by γ11. The constructs y1, y2, and y3 are hypothesized reflectively-measured outcomes of the FMC. Given that x1, x4 and x5 are all exogenous, they are all allowed to co-vary, as indicated by covariances φ14, φ15 and φ45 (MacCallum & Browne, 1993).

Modeling the influence of x1 on y1 along the lines shown in Fig. 1 (i.e., by specifying a direct path between the two constructs only) is intuitively appealing as it simply mimics the way such effects are specified in both CSA and PLS path modeling when a reflectively-measured construct acts as the dependent variable. However, when the latter is an FMC such an approach creates a logical inconsistency because “a change in the value of a formative latent variable cannot occur independently of a change in the value of one or more of its indicators” (Cadogan & Lee, 2013, pp. 234–235). This condition is clearly not fulfilled in the model in Fig. 1 because the antecedent construct (ξ1) impacts the FMC (xi) without impacting any of the formative indicators (x4, x5). Thus, the specification mistakenly assumes that xi completely mediates the impact of ξ1 on y1 and y2 independently from the FMC's formative indicators. As the latter remain exogenous but are allowed to covary both with each other and with ξ1, the direct effect γ11 in fact only captures the antecedent variable's incremental impact conditioned on the formative indicators. Consequently, if the formative measurement model has been correctly specified, no direct influence of ξ1 on y1 will emerge (i.e., within sampling error, γ11 = 0), resulting in a total effect of zero. Given a true non-zero total effect, this specification will considerably underestimate the impact of ξ1 on y1 and invariably lead to wrong conclusions.

Unfortunately, the practice of linking antecedent constructs to a FMC only by direct construct-level paths is widespread in the literature. A review of articles (47 in all) involving endogenous FMCs in six top-tier marketing journals during 2006–2012 (see Table 1) revealed that the overwhelming majority (94%) incorrectly specified the effects of antecedent variables on endogenous FMCs at the construct level only, that is, by directly linking the predictor construct(s) to the FMC. Our review covered the following journals: International Journal of Research in Marketing (IJRM), Journal of the Academy of Marketing Science (JAMS), Journal of Marketing (JM), Journal of Marketing Research (JMR), Journal of Service Research (JSR), and Marketing Science (MS). A full list of the reviewed articles can be found in the Online Appendix A.

As can be seen from Table 1, the reviewed studies have been classified according to (a) whether aggregation of formative indicators was undertaken prior to estimation, and (b) whether the latter was performed with CSA, PLS or regression-based approaches. Both dimensions – handling of formative indicators prior to analysis and estimation method – have a substantial influence on the distortions resulting from incorrectly specifying the links between FMCs and their antecedents. Roughly one-third of the articles used disaggregated formative measurement models, however, in each of these studies the direct ‘construct-level only’ effects specified for the relationships between antecedent constructs and FMCs were interpreted as total effects. More specifically, in CSA, the direct effect was erroneously not regarded as an incremental/additional effect but as the only (i.e., total) effect of the antecedent construct on the FMC. Likewise, in PLS path modeling, direct links have been misinterpreted as representing the total impact of the antecedent construct(s) on the FMC which – as vividly demonstrated by our empirical illustration in Section 4 – is inappropriate. The remaining two-thirds of the articles formed composites prior to CSA or regression-type analysis. Although – at least in principle – this approach allows to correctly estimate total effects (see Section 3.3 for the necessary conditions), the prevailing practice of aggregating equally-weighted formative indicators casts serious doubt on whether such effects have indeed been correctly estimated in the corresponding empirical studies.

Fig. 1. Endogenously-positioned FMC: Misspecified model.
Three papers (Melancon, Noble, & Noble, 2011; Roggeveen, Tsiros, & Grewal, 2012; Seiders, Voss, Godfrey, & Grewal, 2007) are notable exceptions from the mainstream approach in that they at least estimate individual effects of antecedent variables on each of the formative components. However, none of these study documents results for a complete structural model in which the effects of antecedent variables on the FMC are mediated by its formative components. Instead, total effects were determined based on a regression of a composite score for the FMC on the antecedent variables.

In the next section, we provide the correct specification for evaluating effects on FMCs for the three most typical cases: disaggregated formative measures in CSA, disaggregated formative measures in PLS and aggregated formative measures in CSA or regression.

### 3. Correct specification of endogenously-positioned FMCs

#### 3.1. Disaggregated formative measures in CSA

Fig. 2 displays the correct specification in CSA for assessing the impact of the antecedent construct $\xi_i$ on the FMC (now denoted as $\eta_{f1}$). In contrast to the misspecified model in Fig. 1, there are now direct effects (represented by $\gamma_{11}$ and $\gamma_{21}$) from $\xi_i$ on the formative indicators thus making the latter variables endogenous (Temme & Hildebrandt, 2006). This is achieved through the introduction of the pseudo-latent variables $\eta_1$ and $\eta_2$ which enable the formative indicators of $\eta_{f1}$ (now denoted as $y_1$ and $y_2$) to function as endogenous variables; note in this context that $\lambda_{11} = \lambda_{22} = 1$ and $c_1 = c_2 = 0$, which implies that $\eta_1 \equiv \eta_4$ and $\eta_2 \equiv \eta_5$. Similar to the common specification of free covariances between exogenous formative indicators (MacCallum & Browne, 1993), the error terms of the two pseudo-latent variables $\eta_1$ and $\eta_2$ are allowed to correlate (i.e., $\psi_{12} \neq 0$). Following Bollen and Davis (2009), identification of the formative measurement model is established by, first, specifying direct effects from $\eta_{f1}$ on the two reflectively-measured outcome constructs $\eta_4$ and $\eta_5$, and, second, by fixing one of these outgoing paths to unity (i.e., $\beta_{31} = 1$). If we proceed from the full mediation hypothesis implied by a correctly specified formative measurement model (i.e., that the influence of any remote cause has to be completely channeled through the formative indicators), the exogenous latent variable $\xi_i$ impacts $\eta_{f1}$ (i.e., the FMC) only indirectly via its influence on the two pseudo-latent variables $\eta_1$ and $\eta_2$ (model without the dashed path $\gamma_{31}$ in Fig. 2). In this specific case, the total effect of $\xi_i$ on $\eta_{f1}$ equals the sum of the exogenous latent variable’s indirect effects on $\eta_1$ via $\eta_4$ and $\eta_5$, that is, $\gamma_{total} = \gamma_{11}\hat{\beta}_{31} + \gamma_{21}\hat{\beta}_{32}$. Note, in this context, that if, upon estimation of this model a significant

---

**Table 1**  

<table>
<thead>
<tr>
<th>Prior aggregation of formative indicators/components</th>
<th>Estimation method</th>
<th>Linear and non-linear regression, seemingly-unrelated regressions, ANOVA etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No</strong></td>
<td>Collier and Bienenstock (2006b)</td>
<td>Ahearne, MacKenzie, Podsakoff, Mathieu, and Lam (2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Antioco, Monaert, Lindgreen, and Wetzels (2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Davis and Golicic (2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ernst, Hojer, Krafft, and Krieger (2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grégoire, Laufer, and Tripp (2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hennig-Thurau, Henning, and Sattler (2007)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Köhler, Rohm, de Ruyter, and Wetzels (2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>McFarland, Bloodgood, and Payan (2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Miao and Evans (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rapp, Ahearne, Mathieu, and Rapp (2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Storey and Kahn (2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sundaram, Schwarz, Jones, and Chin (2007)</td>
</tr>
<tr>
<td><strong>Yes</strong></td>
<td>Nakata et al. (2011)</td>
<td>Becker, Greve, and Albers (2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fang, Palmatier, Scheer, and Li (2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hornburg, Först, and Koschare (2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kim and Hsieh (2006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lee, Johnson, and Grewal (2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lo, Ghosh, and Lafontaine (2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Menguc and Auh (2006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Menguc and Auh (2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Moos and Gosh (2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Noroff, Kyriakopoulos, Moorman, Paswels, and Dellaeart (2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Piouffe, Hulland, and Wachner (2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Roggeveen et al. (2012)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sarin, Challagalla, and Kohli (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Seiders et al. (2007)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sethi and Iqbal (2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sethi, Iqbal, and Sethi (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stahl, Heitmann, Lehmann, and Nenslin (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Verhoef and Leeflang (2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Voges, Anderson, and Ross (2008)</td>
</tr>
</tbody>
</table>

Notes: We searched all issues in JBRM, JAMS, JM, JMR, JCR and MS published during 2006–2012 for articles including at least one of the following terms: ‘formative’, ‘causal’, or ‘composite’. Next, each article identified in the first step was screened to determine whether it contained at least one endogenously-positioned FMC. Note that we only selected articles where the focal construct(s) were explicitly conceptualized as being formative by the authors (thus, for example, composites build from conventional second-order factor models were excluded). Studies that estimate individual effects of antecedent variables on each formative indicator are indexed a and b, respectively.

* Studies use several aggregation levels and estimation methods.

b Study uses several aggregation levels.
modification index is obtained pointing to an (additional) direct effect from \( \xi_1 \) on \( \eta_3 \), this would be indicative of misspecification of the formative measurement model itself in that at least one relevant formative indicator (correlated with \( \xi_1 \)) has been omitted. Given that, in practice, formative measurement models are sometimes specified somewhat incorrectly, it seems prudent to include an additional direct path to the FMC (model including the dashed path \( \gamma_{31} \) in Fig. 2) which will pick up the antecedent variable's effect on the FMC via any omitted formative indicators. In such an extended model, the total effect amounts to \( \gamma_{total} = \gamma_{11} \beta_{31} + \gamma_{21} \beta_{32} + \gamma_{31} \). Of course, if the direct path turns to be non-significant (i.e., within sampling error, \( \gamma_{31} = 0 \)), it could be subsequently eliminated; indeed, such a non-significant effect would enhance confidence in the FMC's measurement model.

### 3.2. Disaggregated formative measures in PLS

Correctly specifying the same model in PLS (see Fig. 3) leads to a similar model structure as in CSA with two fundamental differences. First, introducing \( \eta_1 \) and \( \eta_2 \) as pseudo-latent variables for the formative indicators, leaves \( \eta_3 \) (i.e., the focal FMC) without indicators. Since such a model cannot be estimated in current implementations of the PLS approach (e.g., SmartPLS or PLS Graph), an additional reflective indicator (shown as \( y_9 \) in Fig. 3) needs to be assigned to \( \eta_3 \). Second, unlike CSA, PLS does not allow disturbance terms to be correlated; thus the covariance between \( \zeta_1 \) and \( \zeta_2 \) is zero (i.e., \( \psi_{12} = 0 \)).

As was the case for CSA, it is recommended to also include a direct path from \( \xi_1 \) to \( \eta_3 \) (shown as the dashed path \( \gamma_{31} \) in Fig. 3) to account for the antecedent construct's impact on the FMC via any omitted formative indicators.
indicators. This is particularly important in PLS as the lack of model fit diagnostics such as modification indexes, can result in misspecifications remaining undetected.

3.3. Aggregated formative measures in regression

As previously noted in Table 1, formative indicators are often aggregated and the resulting composites used as dependent variables in subsequent regression analyses. The key reason for doing so is because “lack of parsimony when modeling formative indicators as separate constructs is an issue” (Howell, Breivik, & Wilcox, 2007, p. 215; see also Cadogan & Lee, 2013). However, there is a clear downside to aggregation in that potential relationships of interest between an antecedent variable and the individual indicators of a FMC cannot be explicitly assessed. Thus one cannot trace the separate paths from the antecedent construct to the FMC via the formative indicators and draw inferences as to which indicators are mostly affected and how. This is an important shortcoming of aggregation not least because it is entirely possible that the same predictor construct can impact some formative indicators positively and others negatively; in this case, aggregating indicators can completely mask the impact of the antecedent construct on the FMC. Furthermore, potentially omitted formative indicators will not be detected if aggregation is used (since there will be only a single path linking the antecedent construct with the composite).

With the above caveats in mind, let us assume that the FMC in Fig. 2 is now represented by the composite variable $C_3$ (instead of the formative latent variable $\eta_3$), completely determined by the two formative indicators $y_1$ and $y_2$ (i.e., $\xi_3 = 0$). Assuming mean-centering of $y_1$ and $y_2$ and given specific, a priori fixed $\beta$-weights, aggregation leads to the following linear composite:

$$C_3 = \beta_{31}y_1 + \beta_{32}y_2.$$  

(1)

For each formative indicator, the impact of the mean-centered antecedent variable $\xi_i$ is specified by the following regression equation:

$$y_i = \gamma_{i1}\xi_i + \xi_i$$  

(2)

where $i = 1, 2$.

The total effect of $\xi_i$ on $C_3$ is then captured by the slope parameter (referred to as $\gamma_{i\text{total}}$) in the corresponding regression of $C_3$ on $\xi_i$ which can be estimated as follows:

$$\gamma_{i\text{total}} = \frac{\text{Cov}(C_3, \xi_i)}{\text{Var}(\xi_i)}$$  

(3)

Substituting Eq. (2) for the two formative indicators in Eq. (3) and invoking the common assumption that the predictor $\xi_i$ is not correlated with the disturbances $\xi_1$ and $\xi_2$ yields:

$$\gamma_{i\text{total}} = \frac{E[(\gamma_{11}\beta_{31} + \gamma_{21}\beta_{32})\xi_i + (\beta_{31}\xi_1 + \beta_{32}\xi_2)|\xi_i]}{\text{Var}(\xi_i)}$$  

(4)

$$= \gamma_{i1}\beta_{31} + \gamma_{i2}\beta_{32}.$$  

Thus, as long as the $\beta$-weights for the formative indicators are correctly determined, regressing $C_3$ on $\xi_i$ yields an unbiased estimate of the total effect of $\xi_i$ on $C_3$ via the formative indicators (note that this conclusion may not apply to the FMC $\eta_3$ if some of the indicators have been omitted). Even in this case, however, the shortcomings of aggregation noted earlier still apply and, therefore, we urge researchers to opt for disaggregated modeling of endogenous formative indicators along the lines discussed in Sections 3.1 and 3.2.

4. Empirical illustration

We now illustrate the steps involved in modeling a structurally endogenous FMC by using real data from a survey of students’ satisfaction with the university cafeteria (n = 181). By comparing the results under both incorrect (i.e., direct effect at the construct level only — Model 1) versus correct specifications (i.e., indirect effects via the formative indicators plus a direct effect at the construct level — Model 2), we empirically highlight the substantial discrepancies in the study implications arising from the two models. Furthermore, we show that CSA and PLS produce vastly different results if the antecedent variable’s effect on the FMC is incorrectly modeled.

Our illustrative model (shown in Fig. 4) proposes that the importance price plays in students’ decisions on where to go for a meal (hereafter simply referred to as ‘price importance’) has an impact on their satisfaction with the university’s cafeteria (hereafter simply called ‘satisfaction’). Satisfaction is conceptualized as a FMC, whereby satisfaction with the (1) taste and (2) healthiness of the meals, (3) the variety of meals offered each day, (4) the cafeteria’s cleanliness, and (5) the prices charged by the cafeteria function as formative indicators. Additionally, satisfaction is measured by two reflective items (overall satisfaction; right choice); all indicators of satisfaction are measured on 5-point rating scales, with larger values implying more favorable judgments. The exogenous antecedent variable price importance is measured by a single item, whereby larger values on a 4-point rating scale indicate higher importance of price. The relevant parameter estimates under both methods of estimation are shown in Table 2.

4.1. Estimation results — Model 1

In Model 1, the antecedent variable price importance impacts satisfaction only directly, that is, the five formative indicators remain exogenous variables. Thus, the corresponding CSA specification drops the paths $\gamma_{11}$ to $\gamma_{15}$, and only models the path $\gamma_{61}$; the formative indicators are allowed to covary freely both with each other and with the antecedent construct, $\xi$. A scale for the FMC was established by appropriately constraining its variance to unity as suggested by Franke, Preacher, and Rigdon (2008), while linking the FMC to the two reflective items ensured that the model was identified (e.g., Bollen & Davis, 2009). In the PLS specification, Mode B was chosen for the outer model in light of satisfaction’s conceptualization as a FMC (e.g., Hair, Hult, Ringle, & Sarstedt, 2013).

CSA estimation of Model 1 provides an excellent overall fit ($\chi^2 = 6.58$, df = 5, $p = 0.254$; RMSEA = 0.042; CFI = 0.995; NNFI = 0.986; SRMR = 0.016). The parameter estimates suggest that the importance price plays in students’ food-related decisions has no significant impact on their satisfaction with the university cafeteria ($\gamma_{61} = 0.03$, $p = 0.62$). In contrast, PLS estimation suggests a strong positive relation between price importance and satisfaction ($\gamma_{61} = 0.31$, $p < 0.0001$). Marked differences also emerge for the formative indicator weights. Whereas in the CSA model all weights are positive and highly significant, only two formative indicators (satisfaction with prices and taste) are significantly linked to satisfaction in the PLS model. Furthermore, price satisfaction has a negative weight which is clearly counterintuitive as it suggests that increasing students’ satisfaction with the prices charged would decrease their overall satisfaction level.

The highly divergent CSA and PLS results for Model 1 can be largely attributed to the fact that CSA and PLS optimize very different criteria during parameter estimation (e.g., see Henseler & Sarstedt, 2013). Whereas CSA minimizes a global fit function based on some distance between the observed and the model-implied covariance matrix, PLS maximizes a correlation-based criterion. Specifically, the PLS solution
is equivalent to the results of a canonical correlation analysis where 'price importance' is the predictor and 'satisfaction' is the dependent canonical variate. Thus, the formative indicator weights $\beta_{61}$ to $\beta_{65}$ are determined such that the correlation (i.e., the PLS estimate for the direct effect $\gamma_{61}$ in Model 1) between the two variates is maximized. As a result, the estimated weights no longer reflect the extent to which the different formative indicators contribute to the overall satisfaction level independently from antecedent variables like price importance.

The above discrepancies between CSA and PLS serve to emphasize that incorrect modeling of endogenously-positioned FMCs may lead to very different conclusions depending on the estimation method used. What is even more important, however, is that neither the CSA nor the

Table 2
Unstandardized parameter estimates for illustrative model.

<table>
<thead>
<tr>
<th>Parameter (from $\rightarrow$ to)</th>
<th>Model 1 Only direct effect on FMC</th>
<th>Model 2 Both indirect and direct effects on FMC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CSA</td>
<td>PLS</td>
</tr>
<tr>
<td>Price importance $\rightarrow$ Satisfaction ($\gamma_{61}$)</td>
<td>0.03</td>
<td>0.31***</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Taste $\rightarrow$ Satisfaction ($\beta_{61}$)</td>
<td>0.51***</td>
<td>0.74**</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>Healthiness $\rightarrow$ Satisfaction ($\beta_{62}$)</td>
<td>0.22***</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.23)</td>
</tr>
<tr>
<td>Variety $\rightarrow$ Satisfaction ($\beta_{63}$)</td>
<td>0.29***</td>
<td>−0.01</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.23)</td>
</tr>
<tr>
<td>Cleanliness $\rightarrow$ Satisfaction ($\beta_{64}$)</td>
<td>0.29***</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.38)</td>
</tr>
<tr>
<td>Prices $\rightarrow$ Satisfaction ($\beta_{65}$)</td>
<td>0.23***</td>
<td>−0.97***</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.19)</td>
</tr>
<tr>
<td>Price importance $\rightarrow$ Taste ($\gamma_{11}$)</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price importance $\rightarrow$ Healthiness ($\gamma_{21}$)</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price importance $\rightarrow$ Variety ($\gamma_{31}$)</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price importance $\rightarrow$ Cleanliness ($\gamma_{41}$)</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price importance $\rightarrow$ Prices ($\gamma_{51}$)</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Standard errors are in parentheses.
* $p < 0.10$ (two-sided test).
** $p < 0.05$ (two-sided test).
*** $p < 0.01$ (two-sided test).
PLS results for Model 1 accurately depict the true influence of price importance on satisfaction. Only Model 2 can unambiguously reveal this influence as shown below.

4.2. Estimation results — Model 2

In Model 2, price importance is supposed to impact satisfaction both indirectly, that is, through the observed formative indicators (which are now specified as endogenous pseudo-latent variables represented by $\eta_1$–$\eta_5$ in Fig. 4) and directly (in order to capture a possible impact via unobserved formative indicators). Since we presume that price importance would not completely account for the observed formative indicators’ intercorrelations, the corresponding error terms (i.e., $\zeta_1$–$\zeta_5$ in Fig. 4) were allowed to covary in the CSA model. In the corresponding PLS model, satisfaction was measured by the two reflective items, that is, Mode A has been chosen for the outer model.

CSA estimation of Model 2 leads to the same overall fit as for Model 1. Likewise, the direct effect of price importance as well as the formative indicator weights are identical to those in Model 1 (see Table 2). However, the results now show that price importance impacts cafeteria satisfaction indirectly through significant positive effects mediated by taste ($p = 0.05$) as well as cleanliness ($p = 0.06$) and through a significant negative effect via prices ($p = 0.03$). These significant but opposing indirect effects ultimately result in a non-significant total effect of 0.13. The latter, however, should not be interpreted as implying that price importance does not affect satisfaction but rather that it has countervailing influences (some positive and some negative) channeled through different formative indicators. In fact, the finding that the same antecedent construct (here price importance) can have significant but directionally opposing effects on different indicators of a FMC (here, satisfaction) corroborates the arguments made in Section 3.3 against the aggregation of formative indicators prior to analysis. Moreover, the non-significant direct path from price importance to satisfaction in Model 2 ($\gamma_{\text{direct}} = 0.03, p = 0.62$) indicates that, at least for the explanatory variable price importance, the satisfaction construct’s content domain seems to be adequately captured by the five formative indicators shown in Fig. 4. Note that the results can now be used to provide specific managerial insights into how satisfaction can be improved. For example, since students that place high importance on price are more satisfied with taste and cleanliness but at the same time are less content...
with the prices charged, communicating a “best deal” or “value for money” image for the cafeteria may help increase students’ satisfaction with the latter.

Unlike with Model 1, PLS estimation of Model 2 largely produces very similar parameter estimates as the CSA analysis (see Table 2; the total effect of 0.12 is non-significant as well) and thus leads to the same substantive conclusions. Thus the correct specification for modeling endogenous FMCs seems to lead to consistent results irrespective of the method of estimation whereas the incorrect specification does not.

5. Guidelines for researchers

Based on our analysis, the following guidelines can be derived regarding the modeling of FMCs in endogenous positions. First, we recommend that researchers avoid aggregation of formative indicators of the endogenous FMC prior to estimation. Instead, a disaggregated approach should be adopted whereby each formative indicator is modeled as a separate endogenous variable and linked to the relevant antecedent construct(s) as in Fig. 2.

Second, we recommend that the influence of antecedent constructs on an FMC should always be modeled by specifying direct effects on the corresponding formative indicators (like the paths $γ_{1}$ and $γ_{2}$ in Fig. 2) instead of direct effects on the FMC only. If the measurement model for the FMC is correctly specified, this should suffice to completely capture the antecedent variables’ impact on the focal FMC since all antecedent variables’ effects on the FMC will be fully mediated by the formative indicators (as was the case in our illustrative example).

Third, we recommend that an additional direct effect (like the path $γ_{3}$ in Fig. 2) is introduced in order to capture any effects of the antecedent variables which are conveyed by unobserved components (i.e., omitted indicators) of the FMC. Testing this effect not only ensures that the total effect of the antecedent variable on the FMC will be correctly estimated but helps scrutinize the validity of the formative measurement model. Since a direct effect of an antecedent variable on an FMC captures its impact over and above that through the formative indicators, a significant direct effect implies that there are other (unobserved) indicators through which the antecedent variable operates. Especially if the direct effect of an antecedent construct on the FMC exceeds (in absolute terms) the corresponding indirect effects via the formative indicators, doubt would inevitably be cast on whether a sufficiently comprehensive set of indicators has been used to operationalize the FMC. In this case, it may be wise to revisit the specification of the formative indicators in light of the construct definition.

Fourth, we recommend that researchers test for such (additional) direct effects after establishment that the proportionality constraints implied by formative measurement models (Bollen & Davis, 2009; Franke et al., 2008) hold for the FMCs in a study. Such constraints emerge because the effects of the formative indicators on downstream outcomes of a FMC (i.e., reflective indicators or dependent constructs) are assumed to be completely channeled through the FMC such that “there are no [additional] direct effects between the indicators and the outcome variables in the model” (Diamantopoulos, 2011, p. 340). To illustrate, if a FMC’s influence on one outcome variable $η_{o}$ is twice as strong as its effect on another outcome variable $η_{e}$, every formative indicator’s effect on $η_{o}$ must likewise be twice as strong as its effect on $η_{e}$ (hence the name ‘proportionality constraints’). If these proportionality constraints do not hold for a particular formative indicator (in CSA such violations can be revealed by inspecting the modification indices for direct effects between formative indicators and downstream outcome variables), this casts doubt on the latter’s proportionality constraints hold before testing for a direct effect of the antecedent variable on the FMC in order to ensure that the latter test is not distorted by inadequate formative indicators.

Fifth, we recommend that researchers opt for CSA rather than PLS when estimating models with endogenous FMCs. This is because (a) PLS requires at least one (additional) reflective indicator if an FMC is endogenous, (b) PLS does not allow the disturbance terms of endogenously-specified formative indicators to be correlated, and (c) PLS offers no diagnostics (such as modification indices) enabling the assessment of proportionality constraints. Taken together, these limitations suggest that CSA should be the method of choice when estimating models with endogenous FMCs.

By means of summarizing the above recommendations, Fig. 5 provides a flowchart that marketing researchers can use for guidance when placing a FMC in an endogenous position in a structural equation model. Extant literature on formative measurement has so far lacked clear and detailed guidelines regarding the proper modeling of endogenously positioned FMCs in structural equations models. Hopefully, the procedures outlined in this paper, go some way toward filling this gap.

Acknowledgment

The authors would like to thank the IJRM Editor, Area Editor, and two anonymous reviewers for helpful comments and suggestions on previous drafts of this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ijresmar.2014.03.002.

References


Full Length Article

Meta-analysis selection bias in marketing research

Martin Eisend *, Farid Tarrahi 1

European University Viadrina, Große Scharrnstr. 59, 15230 Frankfurt (Oder), Germany

1. Introduction

It is a well-known problem that preferential publication of significant and strong results over non-significant and weak results leads to a literature that provides a false impression regarding the size of the effect in question. There is strong evidence from several fields of science that this “publication bias” exists (Dickersin, 2005). By including published and unpublished studies in their quantitative review, meta-analysts try to mitigate the problem that the publication status of a study (i.e., whether the study is published or unpublished) is related to the effect size estimate in the study. The efforts to include studies of various publication statuses and the tendency of meta-analytic authors to select particular studies—whether intentionally or not—are called selection biases (Ferguson & Brannick, 2012). The current study identifies and examines selection bias in 94 meta-analyses in marketing research and its consequences for academia.

Selection bias arises from the selection decision of a meta-analyt, whereas publication bias is based on the decision of authors and editors to submit and to publish a manuscript, which precedes the selection decision of the meta-analyt. Although it is a different kind of bias, a selection bias might have similar consequences as a publication bias because certain studies are more likely to be selected than others, which influences the strength of the meta-analytic estimate and the attention scholars pay to the results. These consequences are of importance for both practitioners and scientists. Biased estimation of effects can lead to wrong decisions of practitioners and cause harm because inefficient measures are chosen. Biased findings can steer future research endeavors and achievements of academics in the wrong direction, lead to wastage (i.e., unnecessary work), and harm the pursuit of scientific truth (Knight, 2003). A thorough investigation of a selection bias is essential to evaluate the true value of meta-analytic findings.

This study contributes to the literature in several ways. First, the study contributes to the research about meta-analyses by examining for the first time the selection bias of meta-analysts and its consequences for academia. Second, the study contributes to our general knowledge about publication bias, which is related to the selection bias. The findings indicate not only that whether a study is published influences the size of an effect (which has been the focus of prior research on publication bias) but also that where (i.e., journal outlet) the study is published can bias the findings reported in the study. Third, the study provides details about the existence and extent of selection bias in the field of marketing. These insights provide implications for marketing researchers on how they should conduct, review, and read current and future meta-analyses.

2. Background and hypotheses

To avoid publication bias, scholars recommend that meta-analysts make a purposeful attempt to collect both published and unpublished studies (e.g., Borenstein, Hedges, Higgins, & Rothstein, 2009). Unpublished studies are produced by academic institutions that
are not controlled by publishers, such as working papers or unpublished doctoral theses (Hopewell, Clarke, & Mallett, 2005).

Meta-analysts have better access to published studies, as a representative sample of unpublished studies does not exist (Ferguson & Brannick, 2012; Kepes, Banks, McDaniel, & Whetzel, 2012). While other scientific fields such as medicine have developed registers that help scientists to track unpublished research, this has not been done in marketing research. It is, therefore, easier and more likely for a meta-analyst in marketing to access and to include primarily published studies compared to unpublished studies. However, even among published studies, those that are published in leading journals are more easily accessible (e.g., the meta-analyst’s academic institution might not have a subscription to non-leading journals).

Meta-analyses vary in the percentage of included studies that have differing publication statuses, which at least partly depends on the efforts a meta-analyst exerts into searching and retrieving the studies (Banks & McDaniel, 2011). For instance, there is meta-analysis that puts much effort into retrieving unpublished studies. Other meta-analyses are based on systematic issue-by-issue searches of particular journals, usually the leading journals in the field as well as topic-related journals. Such issue-by-issue searches increase the likelihood that relevant studies from the searched journals are included. Studies from other journals that are searched by other means (e.g., a keyword search in electronic databases) might be overlooked because effect size estimates worthy of inclusion might not be detected this way (e.g., the relevant effects might not be mentioned in the abstract of the study that is searched for the occurrence of keywords). Furthermore, the number of citations to a study makes it easier to identify a study when searching references of previously found studies, which consequently favors studies published in leading journals that have high citation rates.

The selective sampling of studies can produce an incorrect estimate of the true effect (Renkewitz, Fuchs, & Fiedler, 2011). It is difficult to determine the exact nature of the selection bias, because we do not know the true effect of the relationship that is investigated in a meta-analysis. Whether the findings in top journals are upward biased or the findings in lesser journals or of unpublished studies are downward biased can only be inferred from the empirical distribution of meta-analytic effect sizes (Egger & Smith, 1998).

2.1. The influence of whether and where a study is published on meta-analytic effect sizes

In marketing research, it has been shown that the percentage of significant results reported in journal articles has increased over the years, particularly in the leading journals (Hubbard & Armstrong, 1992). Several studies in medicine and psychology have surveyed reviewers, editors, and authors and found that studies with results rejecting the null hypothesis are more likely to be published (e.g., Coursol & Wagner, 1986; Dickersin, Chan, Chalmers, Sacks, & Smith, 1987; Greenland, 1975). In order to investigate the reasons for the lack of insignificant results in publications, several cohort studies have examined the process from study initiation to dissemination of results by following studies approved by research ethics boards (e.g., Cooper, DeNeve, & Charlton, 1997; Dickersin, 1997; Easterbrook, Berlin, Gopalan, & Menthews, 1991; Olson et al., 2002). They found that the majority of researchers do not submit manuscripts with non-significant results. In addition to the self-selection of authors, the editorial staff is responsible for a publication bias because studies are more likely to be rejected due to the lack of an incremental contribution to the literature.

The publication bias suggests that the publication status is related to the effect size (e.g., Egger, Smith, Schneider, & Minder, 1997; Rust, Lehman, & Farley, 1990). Meta-analytic authors’ tendency to select published studies more than unpublished studies aggravates the publication bias problem (Renkewitz et al., 2011). The more unpublished studies that are included, the weaker the meta-analytic effect size will be.

To date, publication bias studies have focused on the relationship between publication status and effect size by examining whether a study was published. Another plausible, yet barely investigated approach is to search for variations in the quality of publication outlets and their relationship with effect sizes. The underlying idea is that the size of the effect denotes the explanatory potential and, by this, the usefulness of a theory (Aguinis, Dalton, Bosco, Pierce, & Dalton, 2011). The more variance in the dependent variable that is explained, the more useful the underlying theory is thought to be. Combs (2010, p. 11) explains this as follows: “A theory might find support, but its explanatory power—that is, the effect size observed—is so weak that further efforts to develop the theory might not be warranted. … Small effects also raise questions about managerial relevance. … If managers begin to act on theories that are supported by small effects, they are not likely to notice positive results even when they occur.”

Because the standards of methodological rigor and theory development that are considered acceptable in leading journals are higher than in non-leading journals (e.g., Lehmann, 2005; Varadarajan, 2003) and because effect sizes signal a theory’s usefulness and the rigorous application of methods, the editors and reviewers of leading journals are more likely to select studies with strong effect sizes, thus suppressing weak results. Also, authors, who have strong findings or who are more careful and thorough in their work and better control for confounding factors and thus find stronger findings, might be more likely to select these findings for a submission to a leading journal. In other words, censorship due to authors, editors, or reviewers in marketing research is related to the size of effects reported in the studies (Rust et al., 1990).

H1. The ratio of studies published in leading journals in a meta-analysis is positively related to the meta-analytic effect size.

2.2. Consequences of selection bias on publication of and citations to a meta-analysis

Strong and significant effects are considered as important and attract more attention by scholars than weak or non-significant effects. The importance of research findings is evaluated in academia in at least two measurable ways: first, by the gatekeepers of publication outlets (editors and reviewers), who decide which findings are worthy of being published, and second, by scholars who indicate the importance of the findings by citing these studies. Tierney, Clarke, and Stewart (2000) have shown that meta-analyses of individual cancer patient data with significant and impressive results tend to be published in journals with higher impact factors. While the importance of the effect size is rather obvious in medical science, because it indicates how successful treatments and interventions are, studies in business research are more concerned with the mere signification of an empirical finding (Ellis, 2010). We suggest that effect sizes in meta-analyses in marketing research influence their publication success, because they indicate a relative contribution. The magnitude of a meta-analytically derived effect size denotes explanatory potential of theories: theories that explain a larger portion of the variance in relevant outcomes are more useful than those that explain a small portion (Aguinis et al., 2010; Bacharach, 1989). The theoretical relevance increases the likelihood of authors to submit their meta-analysis papers to a top journal and it influences the decision of editors and reviewers to support these papers during the review process. Because we assume that the meta-analytic effect size depends on the publication status of the studies included in the meta-analysis, we formulate the following mediation hypothesis that describes the consequences of selection bias on the probability of a meta-analysis to be published in a leading journal.

H2. The meta-analytic effect size is a mediator for (a) the ratio of unpublished studies and (b) the ratio of studies published in leading journals included in a meta-analysis on the probability that the meta-analysis is published in a leading journal.
Scholars rely on strong effects to use for groundwork in their own research. If meta-analyses are cited based on their effect size, then they are cited according to their relative merit, that is, the ability to explain outcomes of interest (Aguinis et al., 2011). Because meta-analyses do not only focus on the significance of the meta-analytic effect but also on its size, citations are driven by the meta-analytic effect size beyond its mere significance: the stronger the effect, the more likely scholars will refer to the finding and cite it in their own work.

Previous research has shown that journals publishing statistically significant results compared to those publishing null results had on average a higher impact factor, that is, are cited more by scholars (Easterbrook et al., 1991). Meta-analyses that are published in leading journals lead to more citations because these journal vehicles have a higher impact factor. By this, the meta-analytic effect size does not only influence citations directly (i.e., independent on where the meta-analysis is published), it also influences the publication probability of a meta-analysis in a leading journal (Hypothesis H2), which in turn drives citations to the meta-analysis. Therefore, we expect a mediating path next to the direct effect of the meta-analytic effect size on citations.

H3. (a) The meta-analytic effect size increases citations to a meta-analysis. (b) The probability that a meta-analysis is published in a leading journal is a mediator for the relationship between the meta-analytic effect size and citations to a meta-analysis.

Fig. 1 summarizes the relationship between the variables investigated in our study.

3. Method

3.1. Document retrieval

We restrict our search for meta-analyses in marketing research to journals that are currently listed in the social science citation index. To locate meta-analyses in marketing research published by the end of 2012, we searched every marketing journal (see Appendix). The journals that were categorized as marketing journals followed the comprehensive journal classification provided by Harzing’s list (Harzing, 2012). We further include meta-analyses on marketing topics from other business journals that publish marketing studies (e.g., Journal of Applied Psychology and Journal of International Business Studies) and that we found by keyword searches, such as “marketing,” “consumer,” and “meta-analysis,” in EBSCO and Google Scholar.

We define a meta-analysis as the following (see also Richard, Bond, & Stokes-Zoota, 2003): the meta-analysis must report a numerical measure of a relationship between two variables (e.g., correlation or mean difference). The meta-analysis had to systematically summarize evidence of this effect collected within two or more primary scholarly studies (i.e., studies by academic scholars) by two or more researchers or research teams. This definition results in a number of exclusions. We exclude review studies that do not report numerical measures for a relationship between two variables, such as vote-counting studies.

We further exclude studies that compile descriptive results such as the mean survey response rates (e.g., Yu & Cooper, 1983), content analytic percentages (e.g., Abernethy & Franke, 1996), occurrence of analytical approaches (e.g., Dekimpe & Hanssens, 1995), or reliability and validity coefficients (e.g., Homburg, Klarmann, Reimann, & Schilke, 2012). We exclude two review studies that use only databases other than scholarly journals (Lodish, Abraham, Kalmenson, et al., 1995; Lodish, Abraham, Livelsberger, et al., 1995), because it is not possible to decide on a corresponding number of studies nor their publication status. We further exclude studies that meta-analyzed meta-analytic results (e.g., Peterson, 2001). If the meta-analytic data were published in more than one journal article, we refer to the article that was published first. Based on this procedure, we identified 115 meta-analyses in the marketing area.

3.2. Coding

Most meta-analyses provide a single meta-analytic effect size that equals an average of all effect sizes that were provided in the primary studies that were included in the meta-analysis. Because we measure all variables in our models (see Table 1) on the meta-analysis level (that is, the variables vary between meta-analyses), we code for each meta-analysis one mean effect size that integrates all individual effect sizes in the meta-analysis. If the individual effect sizes in a meta-analysis were combined to subgroups of meta-analytic effect sizes rather than into a single meta-analytic effect size (such as, for instance, the effect of humor in advertising on attitudes, intentions, and behavior), the mean effect size was computed by averaging these meta-analytic effect sizes (weighted by the number of underlying individual effect sizes). Most of the meta-analyses in marketing use the correlation coefficient as meta-analytic effect size and therefore we chose the correlation coefficient as the mean effect size, and absolute values were coded because we are interested in the size of the effect, not its direction.

If the meta-analysis applied an effect size metric different from correlation coefficients, we transformed effect sizes to correlation coefficients according to common re-computation methods (e.g., Lipsey & Wilson, 2001). We used the raw mean effect size because this measure was provided in most cases. If the sample size weighted mean or variance-weighted mean was provided, we used this measure because we did not find any difference between raw mean and weighted mean values in meta-analyses that provided both estimates (t = .9, p = .385). Some meta-analyses provided attenuated effect sizes only. Attenuation-corrected estimates are bigger than raw means (t = 4.97, p < .001) and weighted means (t = 4.70, p < .001) because they correct for measurement errors of estimates. The ratio of non-corrected means to attenuation-corrected means that we found in the meta-analyses that provided both estimates is .87. This value is used to correct the estimates from meta-analyses that provide attenuation-corrected estimates only (i.e., we multiply the attenuation-corrected estimates by .87).

![Fig. 1. Relationship between variables investigated.](image-url)
In addition to the effect size, several other dependent, independent and control variables were coded. Table 1 describes these variables and their coding, explains why they were considered (i.e., their function), and provides relevant data characteristics. Because these methodological variables, with the exception of the subject area, are based on codings that did not leave room for interpretation, these variables were double-coded by a second coder for only 25% of the entire conceptual framework, often by means of structural equation models (type 3).

In the final analyses, only 94 meta-analyses were included. Twenty-one meta-analyses were excluded either because the list of manuscripts that were included in the meta-analysis could no longer be retrieved from the authors (eight meta-analyses) or because the meta-analysis does not provide sufficient information to compute a common effect size, because the effect size in the meta-analysis could not be readily transformed into a correlation coefficient (thirteen meta-analyses). That is the case for meta-analysis that apply elasticities (e.g., transformed into a correlation coefficient, because the effect size in the meta-analysis could not be readily calculated from the authors (eight meta-analyses) or because the meta-analysis does not provide sufficient information to compute a common effect size, because the effect size in the meta-analysis could not be readily calculated from the underlying sample size of the primary studies, we receive an average of 55,001 subjects per meta-analysis. A list of all meta-analyses is provided in the appendix.

3.3. Analytical procedure

First, we run different kinds of regression models for each of the dependent variables:

\[ ES_i = \beta_1 + \sum_{i=2}^{16} \beta_i Control_i + \beta_{17} RU_i + \beta_{18} RL_i + \epsilon_i \]  

\[ JO_i = \beta_1 + \sum_{i=2}^{16} \beta_i Control_i + \beta_{17} RU_i + \beta_{18} RL_i + \beta_{19} ES_i + \epsilon_i \]  

\[ CA_i = \beta_1 + \sum_{i=2}^{16} \beta_i Control_i + \beta_{17} RU_i + \beta_{18} RL_i + \beta_{19} ES_i + \beta_{20} JO_i + \epsilon_i \]  

\[ \chi^2 = 5.70, p = .017 \]  

\[ \text{Min.} = .01, \text{max.} = .71, \text{mean} = .27, \text{SD} = .15 \]

\[ \text{Min.} = 0, \text{max.} = .5, \text{mean} = .06, \text{SD} = .10 \]

\[ \text{Min.} = 0, \text{max.} = 1, \text{mean}.39, \text{SD} = .29 \]

\[ 0 = \text{other journal} (71) \quad 1 = \text{leading journal} (23) \]

Notes: IV = independent variable, DV = dependent variable, M = mediator.
The dependent variable effect size (ES, as measured by the Fisher’s z-transformed correlation) is a continuous variable and a linear regression is applied. The dependent variable journal outlet (JO) is dichotomous and we applied a logistic regression. For the variable citations average (CA), we apply a negative binomial regression, because the likelihood ratio test indicates significant overdispersion ($\chi^2 = 117.96; p < .001$). The analysis of the residuals of each model did not indicate any outlier problems nor did we find any other violations of the models.

To test mediation effects, we follow procedures for regression-based mediation tests (Zhao, Lynch, & Chen, 2010). We run three regression models (I to III) that include all control variables and that test: (I) the effect of the independent variable on the dependent variable (c-path), (II) the effect of the independent variable on the mediator variable (a-path), and (III) the effect of both the independent (c′-path) and the mediator variable (b-path) on the dependent variable. Mediation requires at a minimum that the a-path, the b-path and the indirect effect to be significant.

4. Results

4.1. Descriptive results

The mean meta-analytic correlation is .27 (SD = .17). Table 2 presents the correlation matrix of all variables.

4.2. Regression and mediation analysis

Table 3 presents the results of the regression models. The regression model (1) with effect size as the dependent variable shows that the ratio of unpublished studies reduces the meta-analytic effect size ($b = -.36, SE = .20, t = 1.85, p = .069$), supporting the publication bias. The ratio of studies in leading journals enhances the meta-analytic effect size ($b = .15, SE = .07, t = 2.19, p = .031$). The result supports Hypothesis H1.

Fig. 2 presents the results of the mediation analysis (i.e., a-path, b-path, c-path, and c′-path as retrieved from regression analysis). The first diagram shows that there is no direct effect of the ratio of unpublished studies on the probability that the meta-analysis is published in a leading journal. However, the indirect effect is significant as indicated by bootstrapping ($z = 2.66, p = .008$). The non-significant direct effect can be explained by the opposite sign of the indirect effects: the ratio of unpublished studies reduces the meta-analytic effect size that, in turn, increases the probability that the meta-analysis is published in a leading journal. The finding is consistent with the mediation effect that is suggested in Hypothesis H2a.

The second diagram shows that the direct effect of the ratio of studies in leading journals on the probability that the meta-analysis is published in a leading journal is significant. After adding the indirect path, the effect becomes non-significant. The indirect effect is marginally significant as indicated by bootstrapping ($z = 1.96, p = .097$). The finding is consistent with the pattern predicted by Hypothesis H2b.

The third diagram shows that the direct effect of the meta-analytic effect size on citations average is significant ($b = 1.87, SE = .67, t = 2.80, p = .005$), supporting Hypothesis H3a. After adding the mediating path, the effect is reduced but remains significant ($b = 1.49, SE = .27, t = 2.24, p = .025$). The indirect effect is significant as indicated by bootstrapping ($z = 2.75, p = .006$). The finding is consistent with the mediation effect that is suggested by Hypothesis H3b.

The regression results in Table 3 provide some additional findings related to the control variables that deserve attention. As for subject area, we find that meta-analyses dealing with issues related to advertising, channels, and consumer behavior are less likely to be published in leading journals. Meta-analyses dealing with topics related to new product development and strategy receive more citations than other meta-analyses.

As for method differences across meta-analyses, we find that meta-analyses that are mainly based on surveys and field studies receive more citations, but are less likely to be published in leading journals. Furthermore, the meta-analysis type does not influence the meta-analytic effect size, but meta-analyses that include moderators or structural equation models have a higher likelihood to be published in leading journals and receive more citations than meta-analyses that only present and discuss meta-analytic effect sizes.

As for the average publication year of the studies that are included in a meta-analysis, we did not find any influence on effect size, although we were expecting that more recent studies would apply more advanced methods and thus lead to stronger effects. We find a negative effect of the probability to be published in leading journals and on citations average that indicates that meta-analyses on more recent topics have lower publication success.

As for the indicators related to the maturity of the field (number of effect sizes and time frame), we find that the number of effect sizes does not influence the dependent variables, but the time between the oldest and the most recent study that were included in the meta-analysis has a significant effect on the dependent variables. The longer the time frame, the more mature the field and the more established the research questions. At the same time, the less likely the meta-analysis gets published in a leading journal and the fewer citations the meta-analysis receives. This is in line with Rust et al. (1990) who argue that the more mature the field, the more suspiciously reviewers, who are often authors of previous studies, might look at results of additional studies. They might apply more rigorous standards in established research fields and are therefore less likely to accept a meta-analysis for inclusion in leading journals. In a similar way, readers might devote less attention to a meta-analysis once a research field is already fully established. There might be less activity and interest in this area and this lack of interest reduces the publication success.

4.3. Additional analysis

4.3.1. Is variation in publication status indeed driven by the efforts of meta-analysts?

Our central assumption is that the ratio of unpublished studies and the ratio of studies that are published in leading journals in a meta-analysis depend largely on the efforts authors undertake in searching and retrieving studies. An alternative explanation might be that the availability of unpublished or published studies varies over the meta-analyses. To test whether our assumption holds, we additionally coded the methods section in each meta-analysis regarding the efforts authors undertake to retrieve studies with different publication statuses.

Efforts to retrieve unpublished studies go beyond searches that reveal published studies only (e.g., searching in databases with published articles only). Following suggestions in the literature (Hopewell et al., 2005), we distinguish between the following activities of meta-analysts to retrieve unpublished studies: (1) searching databases that include unpublished studies, (2) searching the Internet, (3) calls for unpublished studies (e.g., via ELMARK, (4) contacting authors, and (5) searching dissertations online. Out of 94 meta-analyses, 76 provide sufficient information on the study retrieval process. The number of activities mentioned to retrieve unpublished studies is positively related to the ratio of unpublished studies in the meta-analyses ($r = .60, p < .001$). If we replace the ratio of unpublished studies in regression model (1) with the number of activities to retrieve unpublished studies, we find support for our central assumption, because the number of activities reduces the effect size ($b = -.03, SE = .02, t = 1.83, p = .072$) (see Appendix Table A.4).

We decided to include the number of unpublished studies in our analysis because they are related to the efforts the authors invest in retrieving these studies. Because the percentage of unpublished studies is relatively low, an alternative measure would be one that simply
distinguishes whether unpublished studies were included or not. However, such measure does not as clearly relate to the efforts a meta-analyst invests in retrieval of unpublished studies. For instance, a meta-analyst who includes only one unpublished study out of 100 studies was probably putting less effort in study retrieval than a meta-analyst who includes only one unpublished study out of ten. The dummy variable would indicate in both cases the same value for inclusion of unpublished studies. When including a dummy variable for unpublished studies in regression model 1, the effect is non-significant (b = −.04, SE = .04, t = −90, p = .32). This finding supports the idea that the selection bias is driven by meta-analysts’ efforts during the study retrieval process and not just the fact whether unpublished studies were included, which can have different reasons.

The efforts made to retrieve studies from various journals can best be assessed by the issue-by-issue searches a meta-analyst performs. Only 28 meta-analyses provide information on the journals that were screened in an issue-by-issue search. The ratio of leading journals to all marketing journals that were systematically searched is positively related to the ratio of studies in leading journals that were eventually included in the meta-analysis (r = .60, p < .001). Due to the small sample size of only 28 meta-analyses, we run a regression model where we include only control variables that show a significant

Table 2
Correlation matrix.

<table>
<thead>
<tr>
<th>#</th>
<th>Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Effect size (ES)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Ratio unpublished studies (RU)</td>
<td>−.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Ratio studies in leading journals (RL)</td>
<td>.31</td>
<td>−.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Journal outlet (JO)</td>
<td>.02</td>
<td>−.11</td>
<td>.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Citations average (CA)</td>
<td>.29</td>
<td>.03</td>
<td>.22</td>
<td>.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Control variables

| 6  | Subject area: advertising | .02 | −.06| .12 | −.05| −.14|     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 7  | Subject area: channels    | .09 | −.07| .13 | −.07| −.12|     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 8  | Subject area: consumer behavior | .05| .09 | −.16| −.24| −.07| −.32| −.16|     |     |     |     |     |     |     |     |     |     |     |     |
| 9  | Subject area: methods     | −.05| .16 | −.03| .03 | −.07| −.12| −.08| .01 |     |     |     |     |     |     |     |     |     |     |     |
| 10 | Subject area: new product development | −.10| −.10| −.11| .02 | .09 | −.16| −.10| −.31| −.10|     |     |     |     |     |     |     |     |     |     |
| 11 | Subject area: pricing     | .09 | −.09| .15 | .18 | −.01| −.03| −.09| −.03| −.09| −.11|     |     |     |     |     |     |     |     |     |
| 12 | Subject area: sales       | −.17| .12 | .05 | .07 | −.03| −.14| −.09| −.13| −.09| −.12| −.10|     |     |     |     |     |     |     |     |
| 13 | Subject area: strategy    | .04 | −.09| −.09| .15 | .21 | −.14| −.09| −.28| −.09| −.01| −.10| −.11|     |     |     |     |     |     |     |
| 14 | Meta-analysis type 2      | −.12| .20 | −.05| .01 | −.03| −.08| .09 | .10 | .02 | .12 | −.18| −.03|     |     |     |     |     |     |     |
| 15 | Meta-analysis type 3      | .11 | −.11| .06 | .29 | .15 | .04 | .17 | −.23| −.10| −.02| −.11| .24 | .01 | .53 |     |     |     |     |     |
| 16 | Method type               | .03 | −.14| −.03| .02 | .11 | −.22| .21 | −.43| −.12| .27 | .09 | .17 | .25 | .05 | .04 |     |     |     |     |
| 17 | Year                      | −.11| .08 | −.39| −.23| −.08| −.12| −.04| −.01| .10 | .18 | −.30| −.08| .31 | .06 | .10 | −.01|     |     |     |
| 18 | Number of effect sizes    | .07 | −.09| −.12| .12 | −.01| −.10| −.08| −.06| .02 | .09 | .06 | −.02| −.05| .16 | .05 | .14 | −.11|     |     |
| 19 | Time frame                | −.14| −.07| .11 | −.12| −.11| .11 | .02 | .06 | −.16| −.12| .04 | .24 | .30 | .06 | −.05| .09 | −.60| .12 |     |

Notes: Bold correlations are significant at p < .05, figures in italics are significant at .05 ≤ p < .01 (two-sided tests). Correlations involving dummy variables are (point-)biserial correlations.

Table 3
Model estimation results.

<table>
<thead>
<tr>
<th>#</th>
<th>Variables</th>
<th>(1) Linear regression</th>
<th>(2) Logit regression</th>
<th>(3) Negative binomial regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dependent variable</td>
<td>Effect size</td>
<td>Journal outlet</td>
<td>Citations average</td>
</tr>
</tbody>
</table>

Control variables

|     | Subject area: advertising | .05 (.08)             | −.67 (3.11)*         | −.24 (.44)                      |
|     | Subject area: channels    | .03 (.09)             | −.83 (3.22)*         | −.01 (.44)                      |
|     | Subject area: consumer behavior | .09 (.07) | −.49 (2.59)*         | .34 (.38)                       |
|     | Subject area: methods     | −.01 (.08)            | .82 (2.23)           | −.29 (.44)                      |
|     | Subject area: new product development | −.03 (.08) | 1.31 (1.95)           | .73 (.43)*                      |
|     | Subject area: pricing     | .01 (.07)             | −.03 (2.09)          | .30 (.40)                       |
|     | Subject area: sales       | −.06 (.08)            | 2.09 (2.33)          | .07 (.42)                       |
|     | Subject area: strategy    | .03 (.09)             | 2.04 (2.07)          | .98 (.46)**                     |
|     | Meta-analysis type 2      | −.01 (.05)            | 5.55 (3.29)*         | .56 (.28)**                     |
|     | Meta-analysis type 3      | .07 (.07)             | 8.02 (3.73)**        | .89 (.42)**                     |
|     | Method type               | .05 (.05)             | 3.33 (1.74)*         | .44 (.26)*                      |
|     | Year                      | −.01 (.01)            | −.28 (1.1)*          | −.04 (.16)**                    |
|     | Number of effect sizes    | .01 (.01)             | .01 (.01)            | −.01 (.01)                      |
|     | Time frame                | −.01 (.01)**          | −.14 (.07)*          | −.02 (.01)*                     |

Main variables

|     | Ratio unpublished studies (RU) | −.36 (.20)*          | 2.69 (6.60)          | 3.24 (.99)*                     |
|     | Ratio studies in leading journals (RL) | .15 (.07)**        | 2.38 (2.39)          | .30 (.39)                       |
|     | Effect size (ES)            | 12.69 (4.92)**       | 1.49 (.86)**         |                                |
|     | Journal outlet (JO)         | .27                   | .62                 | .13                             |
|     | R²/Pseudo R²                | 63.35 (17)**         | 64.17 (18)**        |                                |
|     | F (df)                      | 1.79 (16.77)**       |                    |                                |

Unstandardized coefficients and standard errors in brackets are provided.

*p < .10, **p < .05, and ***p < .01 (two-sided tests).
relationship with the main independent and dependent variables (see Appendix Table A.4). In line with our expectations, the ratio of leading journals that were systematically searched increases the effect size ($b = .27, SE = .15, t = 1.84, p = .078$). These results support our assumption that the efforts a meta-analyst undertakes to retrieve different types of studies is related to the percentage of the various studies being included in the meta-analysis, supporting a selection bias.

4.3.2. Are effects in leading journals more likely reported due to size or due to statistical power?

Hypothesis H1 suggests that the size of effect denotes its explanatory potential and is related to the likelihood to be published in a leading journal. An alternative line of reasoning would be that leading journals tend to publish high quality studies with rigorous methods that improve statistical power. Hence, when studies are selected on the basis of statistical significance (and not based on the size of the effect), leading journals also include studies with relatively small effect sizes (significant due to the rigorous methods), which leads to predictions opposite to Hypothesis H1. Two major drivers of statistical power are sample size and reliability of measures. More reliable measures increase statistical power and so does sample size. That is, as sample size and reliability go up, p-values go down and smaller effects pass the significance threshold.

However, we did not find any relationship between the ratio of top journals included in a meta-analysis and the average sample size per effect size ($r = -.09, p = .797$, for $n = 11$ independent effect sizes; $r = -.19, p = .18$, for $n = 49$ dependent effect sizes). We did not find any relationship with the reliability of measures either, as indicated by the ratio of non-corrected means to attenuation-corrected means in a meta-analysis ($r = -.35, p = .18, n = 16$). Due to the small number of meta-analyses for these measures, we included these measures in model 1 only with the main variables (none of the control variables except for number of effect sizes was correlated with the additional variables). The results remained unchanged. The finding indicates that leading journals indeed report stronger effect sizes and not just effects with higher statistical power.

4.3.3. Does this meta-meta-analysis itself have a selection bias?

This study is restricted to meta-analyses that were published in journals that are listed in the SSCI database, because we are interested in investigating the citation outcomes of a meta-analysis. The selection of meta-analyses might have caused a selection bias itself, because the meta-analyses included in the study and in particular their findings might not be representative. That is, there might be additional meta-analyses that are not included in the SSCI database and that report meta-analytic effect sizes that are smaller than the ones reported in our study.

To check such a selection bias, we searched the literature for meta-analyses that were not (yet) published by 2012. We found nine meta-analyses that were either not published or published as extended abstracts only (see Appendix Table A.5). Although the unpublished meta-analyses tend to show smaller meta-analytic effect sizes than the published meta-analyses ($-.19$ vs. $.27$), the difference is not significant ($t = 1.41, p = .16$) which can be due to the small number of only nine unpublished meta-analyses. More importantly, however, when applying regression model (1) to the extended sample of meta-analyses, the results remain stable: the ratio of unpublished studies reduces ($b = -.37, SE = .19, t = 1.97, p = .052$), and the ratio of studies in leading journals increases the effect size ($b = .15, SE = .07, t = 2.15, p = .034$). That is, the influence of a selection bias of the studies included in a meta-analysis does not depend on a possible selection bias of the meta-analyses that we choose for our study.

We further apply the trim and fill method that checks whether the meta-analyses in our sample provide meta-analytic effect sizes that follow the symmetry assumption of the funnel plot, that is, the graphical distribution of the meta-analytic effect sizes and their variances (Duval...
& Tweedie, 2000). In case of a selection bias, this distribution would be asymmetrical and reveal that meta-analyses with small meta-analytic effect sizes are underrepresented. The trim and fill method shows that the distribution of meta-analytic effect sizes is symmetrical and does not show any gaps. The symmetry of the funnel plot indicates that we can assume that the mean value of the meta-analytic correlations of .27 is a good estimate of the true value and that an increasing ratio of unpublished studies leads to a downward bias of effect sizes, while an increasing ratio of published studies leads to an upward bias of effect sizes.

4.3.4. Are the controls for subject area too coarse?
Each meta-analysis covers a different topic and the eight subject areas we applied as control variables might be too coarse and inaccurate to capture substantive differences between the meta-analyses and their findings. While we cannot control for each meta-analytic topic, we applied a finer categorization as suggested by Stremersch, Verniers, and Verhoeof (2007), who distinguish between 19 different subject areas. When using these finer categories of subject areas in our regression models, the findings still support our hypotheses and only the effect of the ratio of unpublished studies on the effect size (regression model 1) becomes weaker and significant only when assuming a one-sided test ($b = -.30, SE = .19, t = 1.54, p = .064$).

5. Discussion
The findings of this study show that whether and where a study that is included in a meta-analysis is published influence meta-analytic effect size estimates and that such a selection bias increases the probability of a meta-analysis to be published in leading marketing journals as well as the number of citations the meta-analysis receives.

First, this study contributes to research about meta-analyses by examining the selection bias of meta-analysts and its consequences for academia. The meta-analytic effect size influences both the attention and evaluation outcome of editors, reviewers, and academic readers, as indicated by the likelihood to be published in leading journals and the citations to these meta-analyses. Selection bias can drive meta-analytic effect sizes upward and fosters these publication outcomes when unpublished studies are neglected and studies in leading journals are preferred by the meta-analyst. Meta-analyses are cited at a significantly higher rate than primary-level studies (Aguinis et al., 2011) becomes weaker and significant only when assuming a one-sided test ($b = -.30, SE = .19, t = 1.54, p = .064$).

Second, the study contributes to our general knowledge about publication bias. The findings show not only that whether a study is published influences the size of an effect (which has been the focus of prior research on publication bias) but also that where the study is published (i.e., in leading journals) can bias the findings reported in the study. Notably, the results related to studies from leading journals are more indicative of a bias, because they reflect the final selection outcome of a review process. Unpublished studies, on the other hand, might be published at a later point and are consequently a less precise predictor of a selection bias.

Third, the study provides details about the existence and extent of selection bias in the field of marketing research. On average, a meta-analysis includes three unpublished studies. The regression coefficients show that increasing the number of unpublished studies by 25% (that is from three to four studies), would reduce the average meta-analytic correlation from .27 to .24. The percentage of studies from leading journals compared to studies from all journals listed in the SSCI is almost 40%. If this number would decrease to 20%, the effect size would decrease from .27 to .25. Both differences are significant ($p < .05$) for samples of 8000 subjects, respectively 4000 subjects and more. This sample size is not uncommon in meta-analyses; less than 15% or the 60 meta-analyses in our sample that provide information on the sample size are based on an overall sample that is smaller than 4000 subjects. Although the selection bias can affect the meta-analytic findings, the difference of .02 or .03 is small. That is, the effect of the bias has only a small impact on overall effect size, and this probably means in practice that what we currently hold to be “true” with respect to the conclusions drawn from existing published meta-analyses is actually still “true”. Nevertheless, knowing the effect of a selection bias can help to achieve more accurate findings in meta-analyses.

5.1. Implications
To achieve more accurate findings in meta-analyses, editors, reviewers, and the academic community in marketing should (1) improve the reporting of a possible selection bias and (2) try to avoid a selection bias.

5.1.1. Improve the reporting of a possible selection bias
During the coding process for this study it became apparent that the documentation of meta-analysis methods in marketing research barely follows any standards and can be considered poor compared to reporting in other areas. For instance, 27.8% of the meta-analyses (32 out of 115) did not include the list of primary studies. In management research, Aguinis et al. (2011) report that only 7.6% of the meta-analyses did not provide such a list. Because methodological decisions in meta-analysis can influence meta-analytic outcomes (Wanous, Sullivan, & Malinak, 1989), it is important to provide as much information as possible on these decisions and to follow common reporting standards. Reporting standards have not been established in marketing research, but in neighboring fields such as psychology (APA, 2010) or economics (MAER Network, 2013). A successful example for such standards is found in medical science, where several scholars have developed the PRISMA statement (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009). The statement aims to help authors improve the reporting of meta-analyses by offering a flow diagram on the different phases of a meta-analysis and a checklist of 27 items pertaining to the content of meta-analysis. Many journals publishing health research nowadays refer to PRISMA in their instructions to authors. Among other things, the checklist includes items that advise meta-analysts how and what to report as related to the review protocol, study eligibility criteria, information sources, search procedure, and the study selection process. If journal editors adopt these guidelines in marketing research, they can support authors who are in the process of writing up a meta-analysis, and they can help reviewers, editors, and readers in evaluating the meta-analysis not just with respect to a possible selection bias. The major academic associations in marketing should support the installation of such guidelines by agreeing on a common standard that is adopted by all journals.

In addition, meta-analysts are well-advised to address the problem of selection bias by considering the influence the publication status has on the meta-analytic effect size. This can be accomplished by including moderator variables for publication status such as top journal vs. others or by using the impact factor of journals. Of the 115 meta-analyses we identified, only eight meta-analyses (7%) included a moderator variable that tests for differences of effect sizes across different publication outlets, with six out of eight meta-analyses providing significant differences that support the findings of our study.

5.1.2. Avoiding a selection bias
The main task to avoid a selection bias for the meta-analysts is a thorough systematic search of the literature. The findings of this study show that this is not done consistently in meta-analyses in marketing research. Too often, the literature search is limited to a few electronic databases. Rothstein (2012) provides an excellent overview and description of a rigorous and thorough literature search, that helps to minimize the potential for selection bias. With the increasing availability
of unpublished work on the internet (e.g., via Google Scholar), the search process becomes less time-consuming and more efficient. In addition, the findings in our study suggest that postings in newsgroups and personal mails to researchers seem a fruitful way to access unpublished work. The effects of these efforts can be inferred from our findings: Performing only one additional research strategy reduces the average meta-analytic effect size from .27 to .26. Adding more non-leading journals to the issue-by-issue search such that the ratio of leading journals drops from an average of 50% to 25% would reduce the effect size from .27 to .25.

Another solution that requires a concerted effort of journal editors and academic associations in marketing is the installation of research registers and outlets for studies with null results. Research registers have been developed in medical science, where each research project is registered at its inception. Many top-tier journals in the medical sciences do not publish studies unless their samples were registered prior to completion of the study (Laine et al., 2007). Such registers provide information about research projects that get published and that remain unpublished and allow a more systematic literature review for meta-analysis. To make sure that all research projects in marketing are registered at its inception, journal editors need to agree to publish studies only if they have been registered upfront. An additional option that makes it easier for meta-analysts to retrieve and include non-significant results and small effect sizes in their meta-analysis is the implementation of outlets (i.e., journals or journal sections) that publish studies with null results. Such non-significant results are not only informative for meta-analysts, they can help scholars in preventing them from investments in uninteresting research questions. Table 4 summarizes the recommendations for meta-analysts, reviewers, editors, and academic associations on how to improve the reporting of a selection bias and how to avoid such a bias.

### 5.2. Limitations and further research

One limitation refers to the concept of unpublished studies as it is commonly used in the meta-analytic literature: Unpublished studies are operationalized as studies that are available but not published. Beyond these studies, researchers might even refrain from writing up manuscripts for projects with non-significant results. As long as research registers have not been developed, such non-significant results might largely remain undetected, increasing both publication and selection bias. Such results can only be retrieved by addressing researchers directly, which has been done by a few meta-analysts in our study. Furthermore, some non-published studies will be published in a journal later, while some will never be published in a journal. However, even the later publication does not necessarily undermine the problem and the influence of publication status on effect sizes: it frequently occurs that only the significant and important findings will make it to the journal while the less important and non-significant findings do not get reported at all (Chan, Hróbjartsson, Haahr, Gotzsche, & Altman, 2004). The reasons are that authors decide not to include certain results when submitting a study to a journal or editors ask to remove specific findings because they are deemed not interesting or simply to save journal space (Banks & McDaniel, 2011).

Another limitation is that we cannot fully determine the true value of the meta-analytic effect size, but a value that is based on an empirical distribution that itself underlies a sampling error. Hence, the figures we present are relative figures and they show how much the empirical effect size changes when different types of studies are included and how much the meta-analytical include this effect size by applying the suggested strategies.

### 5.3. Conclusions

This study examined the issue of selection bias in meta-analyses in marketing research by looking at meta-analytic effect sizes and their drivers and consequences. These effect sizes depend on whether and where a study included in a meta-analysis is published. The meta-analytic effect sizes steer the attention and the evaluation of a meta-analysis by other scholars. The main conclusion of the findings is that meta-analysts, reviewers, editors and the academic community should improve the reporting of a selection bias and—if possible—avoid such bias, because it is in the field’s interest, like in any science, to concern itself with unbiased estimates and accurate findings. This helps to support the merit of meta-analyses that are both influential and important tools in research, because they develop empirical generalizations and summarize knowledge in an area.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ijresmar.2014.03.006. Estimation code for this article can be found online at http://www.runmycode.org. Interested scholars may contact either the corresponding author or IRJM’s editorial office in order to request the dataset.

### References


Sampling, discounts or pay-what-you-want: Two field experiments

Ju-Young Kim a,⁎, Martin Natter a,1, Martin Spann b,2

a Goethe-University Frankfurt, Department of Marketing, Gruenewergplatz 1, 60323 Frankfurt am Main, Germany
b Ludwig-Maximilians-University Munich, Institute of Electronic Commerce and Digital Markets, Geschwister-Scholl-Platz 1, 80539 Munich, Germany

1 Introduction

A considerable and increasing portion of marketing spending is invested in sales promotions, such as price discounts or sampling (Kotler, 2003). According to the IEG Report (2012), US spending on sales promotions grew by 1.7% from 2010 to 2011. In Europe, the price promotion share of revenues in total FMCG revenues by country is approximately 16–30%, and it continues to grow (GK & SAP Study, 2011).

Using sales promotions, companies are able to effectively influence consumers’ short-term and long-term buying behaviors (Bawa & Shoemaker, 2004; Lammers, 1991). On a short-term basis, typical price promotions such as price discounts help increase sales via purchase acceleration (Blattberg, Eppen, & Lieberman, 1981; Jain & Vilcassim, 1991), brand and/or store switching (Aliawadi, Gedenk & Neslin, 1999; Bucklin, Gupta, & Siddarth, 1998; Heerde, Gupta, & Wittink, 2003; Kumar & Leone, 1988) or additional consumption (Aliawadi & Neslin, 1998; Nijs, Dekimpe, Steenkamp, & Hanssens, 2001). However, short-term effects of price promotions are found to die out in subsequent weeks or months (Pauwels, Hanssens, & Siddarth, 2002; Srinivasan, Popkowski Leszczyc, & Bass, 2000) for both manufacturers and retailers (Srinivasan, Pauwels, Hanssens, & Dekimpe, 2004). The authors investigate pay-what-you-want (PWYW) as an alternative promotional tool to free sampling and price discounts in two field experiments. The authors find significant differences in perceived promotional characteristics and relevant performance measures, such as trial and repeat purchases by new customers. The entertaining and innovative character of PWYW induces many people to try it. PWYW may yield a higher repeat purchase rate of new customers, and sellers using PWYW benefit from higher word-of-mouth behavior. Finally, PWYW yields the highest promotional revenues.

Non-price promotions, such as sampling, may be effective in increasing sales over an extended period of time (Bawa & Shoemaker, 2004). Free samples may affect image perceptions and attitudes (Amor & Guilbert, 2007; Hamm, Perry, & Wynn, 1969; Motes & Woodside, 2001), improve brand loyalty (Gedenk & Neslin, 1999) and increase word-of-mouth (WOM) advertising (Holmes & Lett, 1977). In contrast to price discounts, free samples are most often used to stimulate the trial of a new or improved brand; however, such samples may also help to encourage new uses for an established brand or may attract consumers who have just entered the product category (Schultz, Robinson, & Petrin, 1998).

Companies are continuously seeking opportunities to reach new customers more effectively (Promo Sourcebook, 2010). In this study, we investigate a new sales promotion mechanism that shares some characteristics with price promotions and some characteristics with non-price promotions, such as sampling. Pay-what-you-want (PWYW) is a pricing mechanism that allows consumers to pay any price, including zero. If consumers decide to pay a price lower than the regular price, PWYW resembles price discounts with the difference being that consumers determine the discounts. At the same time, PWYW entails elements of sampling because consumers may self-align their prices (even down to a price of zero), which may reduce consumers’ purchase risk for the new product — inducing trial. Due to the possibility of setting a price of zero, PWYW has the potential to penetrate the market in a manner similar to that of sampling (Chen, Koenigsberg, & Zhang, 2010). However, PWYW may also help to screen out consumers who are truly disinterested in the product because of the mental effort of determining an appropriate price and the payment itself. Therefore, PWYW may function as an intermediate instrument between traditional price promotions, such as discounts, and non-price promotions, such as sampling.
The goal of this study is to analyze the effects of PWYW as a promotional tool with respect to its ability, compared with sampling and price discounts, to attract consumers to participate in a promotion. We further investigate repeat purchases of new customers. We test the relative effectiveness and perceptions of PWYW and free sampling in two different field experiments involving Gillette razors (a FMCG product) and photo portrait prints (which is primarily a service). In the second field experiment, we expand the comparison of the promotion types to price discounts. Finally, we derive implications for the favorable usage of PWYW.

In the following section, we summarize the literature on PWYW. In Section 3, we discuss the attractiveness of PWYW, sampling and price discounts with respect to their perceived promotional characteristics, and we formulate research questions and expectations. In Section 4, we outline our data collection and measures and discuss the results from our field experiments. We conclude the paper with a general discussion of implications and a presentation of limitations and future research.

2. Previous literature on pay-what-you-want pricing

PWYW is defined as “a participative pricing mechanism that delegates the whole price determination to the buyer. The seller simply offers one or more products under PWYW conditions, whereas the buyer decides on the price” (Kim, Natter, & Spann, 2009). In practice, PWYW is applied both as a long-term pricing tool (e.g., Kish, Frankfurt or Wiener Deewan, Vienna) and for promotional activities (e.g., Radiohead, Humble Bundle). In addition to music downloads or video games, PWYW is most often applied in service industries, such as restaurants (e.g., Sobo Bistro, Sydney) and hotels (e.g., IBIS Singapore).

Research on PWYW pricing remains scarce. Previous researchers have analyzed the effects of PWYW on sales and revenues and attempted to explain payment motives (Gneezy, Gneezy, Nelson, & Brown, 2010; Kim, Kaufmann, & Stegeman, in press; Kim et al., 2009; León, Noguera, & Tena-Sánchez, 2012; Riener & Traxler, 2012). Kim et al. (2009) argue that, based on the theory of exchange relationships (Heyman & Ariely, 2004), paying nothing may result in distress and social disapproval by others because the relationship between buyer and seller is less governed by market exchange norms and more by social norms, such as norms of distribution or norms of reciprocity. Schmidt, Spann, and Zeithammer (2014) show that theories of pro-social behavior can explain buyer behavior in PWYW.

Recent studies on PWYW repeatedly show that the price paid is significantly different from zero (Kim et al., 2009; Riener & Traxler, 2012; Regner & Barria, 2009; Gneezy, Gneezy, Riener, & Nelson, 2012; Kim et al., 2013). Some evidence also suggests that sellers may use PWYW as a pricing instrument even in the long run, as prices do not decrease over time (Kim, Natter, & Spann, 2010; Riener & Traxler, 2012).

Gneezy et al. (2010) showed that comparing with a fixed price with a charitable component, combining PWYW pricing with charitable donations can increase the sales volume and profit per product. Additionally, Johnson and Cui (2013) and Kim et al. (2013) found that providing external reference prices is advantageous for the seller because people use this information as an anchor in the price-setting process. In addition, the personal interaction between a seller and a buyer, as opposed to an anonymous purchase on the Internet, appears to benefit the seller because the price paid increases (Kim et al., 2013).

In general, PWYW is primarily used for products with low variable costs (e.g., services and digital products) because the risk of selling the product below cost is high.

3. Promotional characteristics and effects of PWYW vs. sampling and discounts

In this section, we discuss our research questions and expectations with respect to the attractiveness of PWYW as a promotional tool compared with sampling and price discounts. In particular, we address differences in perceived promotional characteristics as well as their effects on trial and repeat purchases by new customers.

3.1. Perceived promotional characteristics

Depending on perceived promotional characteristics as well as individual consumer characteristics, consumers decide whether they wish to participate in one specific promotion, thereby balancing the pros and cons of participation. Historically, most behavioral research on promotions have explained consumers’ responses to price promotions in terms of the monetary savings they receive (Blattberg & Neslin, 1993). However, according to Chandon, Wansink, and Laurent (2000), consumers also respond to sales promotions because of additional utilitarian and hedonic benefits, such as the need for entertainment and the need for innovativeness, as well as the perceived effort that the promotion may provide. To assess the anticipated consumer reaction on PWYW promotions compared with sampling and discounts, we need to understand the perceived promotional characteristics. Therefore, our first research question is:

RQ1: What are the perceived promotional characteristics of PWYW compared with sampling and discounts with respect to monetary benefit, entertainment, innovativeness and effort?

Comparing the perceived monetary savings, we expect consumers to perceive the highest monetary benefit in the sampling promotion because free samples are completely free of charge. In contrast, consumers may perceive (fixed) price discounts to be less beneficial compared with sampling and PWYW. With PWYW, consumers may align their prices with their experience or their expectations about the product. However, with PWYW, consumers are asked to make a decision on the price they want to pay, whereas free samples represent a potentially higher monetary benefit for the consumer. Therefore, monetary savings may be lower with PWYW compared with sampling but higher compared with price discounts.

Anticipating consumers’ needs for entertainment, sampling may be an adequate promotional alternative. Previous research has shown that sampling is beneficial for consumers and may be a pleasant experience (Herzog, Heiman, McWilliams, Shen, & Zilberman, 2001; Shampianer, Mazar, & Ariely, 2007). Compared with common price promotions, such as price discounts, Chandon et al. (2000) also showed that non-price promotions, such as sampling scored higher in terms of entertainment. Free products may lead to an overreaction. Shampianer et al. (2007) found that consumers evaluate free products as being beneficial but not solely in terms of the decreased costs; free offers also evoke more positive affective responses. The latter finding implies that with PWYW, when there is an actual choice to pay nothing, positive affective responses may occur. Evidence from previous research suggests that PWYW is at least a popular promotional tool: in Kim et al. (2009), 87.3% of the buyers stated their preference for PWYW over a fixed price program. The self-determination of prices may also be an entertaining component. Consequently, we conclude that sampling and PWYW feature higher entertainment components than price discounts.

Regarding innovativeness, we expect PWYW to achieve the best evaluation. For many years, companies have been using sales promotions such as price discounts and free samples. Most shoppers already have experience with these two promotional instruments. In contrast, PWYW is relatively new and surprising for many people. The press and media have reported on recent PWYW cases (e.g., Mindlin, 2009, in The New York Times). Because new promotions may fulfill an intrinsic need for exploration (Baumgartner & Steenkamp, 1996), we assume that PWYW induces the highest score on innovativeness.

If promotion participants are not required to complete a form (e.g., their contact information), both sampling and price discounts are not associated with any substantial effort. With PWYW, consumers must also decide on the price paid. This decision may be considered

\[
\text{Q1: What are the perceived promotional characteristics of PWYW compared with sampling and discounts?}
\]

It depends on perceived promotional characteristics as well as individual consumer characteristics. Consumers decide whether they wish to participate in one specific promotion, thereby balancing the pros and cons of participation. Historically, most behavioral research on promotions have explained consumers' responses to price promotions in terms of the monetary savings they receive (Blattberg & Neslin, 1993). However, according to Chandon, Wansink, and Laurent (2000), consumers also respond to sales promotions because of additional utilitarian and hedonic benefits, such as the need for entertainment and the need for innovativeness, as well as the perceived effort that the promotion may provide. To assess the anticipated consumer reaction on PWYW promotions compared with sampling and discounts, we need to understand the perceived promotional characteristics. Therefore, our first research question is:

RQ1: What are the perceived promotional characteristics of PWYW compared with sampling and discounts with respect to monetary benefit, entertainment, innovativeness and effort?

Comparing the perceived monetary savings, we expect consumers to perceive the highest monetary benefit in the sampling promotion because free samples are completely free of charge. In contrast, consumers may perceive (fixed) price discounts to be less beneficial compared with sampling and PWYW. With PWYW, consumers may align their prices with their experience or their expectations about the product. However, with PWYW, consumers are asked to make a decision on the price they want to pay, whereas free samples represent a potentially higher monetary benefit for the consumer. Therefore, monetary savings may be lower with PWYW compared with sampling but higher compared with price discounts.

Anticipating consumers' needs for entertainment, sampling may be an adequate promotional alternative. Previous research has shown that sampling is beneficial for consumers and may be a pleasant experience (Heiman, McWilliams, Shen, & Zilberman, 2001; Shampianer, Mazar, & Ariely, 2007). Compared with common price promotions, such as price discounts, Chandon et al. (2000) also showed that non-price promotions, such as sampling scored higher in terms of entertainment. Free products may lead to an overreaction. Shampianer et al. (2007) found that consumers evaluate free products as being beneficial but not solely in terms of the decreased costs; free offers also evoke more positive affective responses. The latter finding implies that with PWYW, when there is an actual choice to pay nothing, positive affective responses may occur. Evidence from previous research suggests that PWYW is at least a popular promotional tool: in Kim et al. (2009), 87.3% of the buyers stated their preference for PWYW over a fixed price program. The self-determination of prices may also be an entertaining component. Consequently, we conclude that sampling and PWYW feature higher entertainment components than price discounts.

Regarding innovativeness, we expect PWYW to achieve the best evaluation. For many years, companies have been using sales promotions such as price discounts and free samples. Most shoppers already have experience with these two promotional instruments. In contrast, PWYW is relatively new and surprising for many people. The press and media have reported on recent PWYW cases (e.g., Mindlin, 2009, in The New York Times). Because new promotions may fulfill an intrinsic need for exploration (Baumgartner & Steenkamp, 1996), we assume that PWYW induces the highest score on innovativeness.

If promotion participants are not required to complete a form (e.g., their contact information), both sampling and price discounts are not associated with any substantial effort. With PWYW, consumers must also decide on the price paid. This decision may be considered

\[
\text{Q1: What are the perceived promotional characteristics of PWYW compared with sampling and discounts?}
\]
carefully because, on the one hand, monetary benefit may be gained from the self-determined discount. However, on the other hand, there is the possible distress of having paid too little, thus violating social exchange norms (Ariely, Bracha, & Meier, 2009; Elster, 1989; Venkatesan, 1966). Thus we expect that PWYW is associated with higher perceived effort than sampling or price discounts.

3.2 Trial and non-promotional repeat purchases of new customers

Perceived promotional characteristics are expected to affect consumer trials. A higher perceived benefit renders it more likely that a consumer will participate in the promotion. Hence, trial represents our first performance measure. Further, the type of promotion may not directly affect repeat (non-promotional) purchases of new customers but may be an indirect link because repeat purchases result from trial. Thus, repeat (non-promotional) purchases of new customers serve as our second performance measure. Our second research question relates to the differential performances that are related to these three promotional tools:

RQ2: What is the performance of PWYW compared with sampling and discounts with respect to trial and repeat (non-promotional) purchases of new customers?

4. Field experiments

In the following, we present two field experiments in which we investigate our two research questions. First, we tested PWYW vs. free sampling for men’s razors of a well-known brand, which are common products to sample. In our second field experiment, we tested PWYW vs. free sampling and price discounts for portrait prints at photographic studios.

Table 1
Field Experiment 1: Selected questions from the surveys.

<table>
<thead>
<tr>
<th>Question</th>
<th>PWYW</th>
<th>Free sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you used the razor yet?</td>
<td>Yes/no</td>
<td>Yes/no</td>
</tr>
<tr>
<td>Do you still use the razor?</td>
<td>Yes/no</td>
<td>Yes/no</td>
</tr>
<tr>
<td>Have you already bought corresponding razor blades?</td>
<td>Yes/no</td>
<td>Yes/no</td>
</tr>
<tr>
<td>How did the prospective savings affect your decision to participate?</td>
<td>1 = very much; 5 = very little</td>
<td>1 = very much; 5 = very little</td>
</tr>
<tr>
<td>This promotion is fun.</td>
<td>1 = very innovative; 5 = not innovative at all</td>
<td>1 = very innovative; 5 = not innovative at all</td>
</tr>
<tr>
<td>This promotion is interesting.</td>
<td>1 = very difficult; 5 = not difficult at all</td>
<td>1 = very difficult; 5 = not difficult at all</td>
</tr>
<tr>
<td>How much are you generally interested in razors?</td>
<td>1 = very interested; 5 = not interested at all</td>
<td>1 = very interested; 5 = not interested at all</td>
</tr>
</tbody>
</table>

The complete survey can be found in Appendix A.

4.1 Field Experiment 1

4.1.1 Data collection and measures

In our first field experiment, we tested the effectiveness of PWYW and free sampling in two comparable hyperstores in the Frankfurt (Germany) metropolitan area. For two consecutive days (Friday and Saturday), we simultaneously distributed Gillette Fusion razors for free (sampling) in one store and asked consumers to pay what they wanted for the razors (PWYW) in the other store. The two hyperstores are part of the same trade chain and are comparable in terms of their location, size, assortment, purchasing power and sales. The distance between the stores is over 50 km (31 miles). Therefore, it is unlikely that the same customers visited both stores on the same weekend (with a few exceptions, stores are closed on Sundays in Germany). The promotion was advertised by A-boards positioned in the checkout area during the two days. In both stores, two professional salespersons who were located near the exit distributed items either for free (S1, store 1) or under PWYW conditions (S2, store 2). Gillette Fusion razors were distributed in this manner during the promotion days. At the PWYW store (S2), an additional salesperson was responsible for the moneybox. All participating customers (S1 and S2) were required to complete a short questionnaire and to answer questions referring to their level of experience with the brand and the product category. An overview of the questionnaires is available in Appendix A.

To analyze the participants’ perceptions of the promotions and how the promotions affected trial and repeat rates among new customers, all participating customers left their phone numbers for the post-promotion survey. After completing the questionnaire, customers were given the razors. At S2, surveyed customers also paid their self-determined price before receiving the product. A total of 625 (S1: 289, S2: 336) customers participated in the survey. Furthermore, we ensured that no out-of-stock situations occurred at either location.

After five weeks, we began our first telephone survey (see Table 1) and for two weeks we contacted the participants of the sampling and the PWYW treatment (in random order). We asked the participants, if applicable, to evaluate their experience with the product and whether they had already bought corresponding razor blades after the promotion.

We asked additional questions to measure the participants’ attitude toward the promotion and the seller. The responses to these questions were measured on a five-point Likert scale (see Appendix A). One year later, we conducted a second telephone survey to test the robustness of our findings. Again, we contacted all the participants in the sample and PWYW treatments and asked whether they still use the product and whether they had bought any Gillette Fusion razor blades. Furthermore, we asked how much interest the participants generally had in razors and to what extent the prospective savings of the promotion affected their decision to participate. The participants further ranked the innovativeness of the promotion and the effort involved with the PWYW promotion. Again, we measured the responses to these questions on a five-point Likert scale.
Table 3
Field Experiment 1: Results.

<table>
<thead>
<tr>
<th></th>
<th>PWYW</th>
<th>Sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trial</strong></td>
<td>336</td>
<td>289 n.s.</td>
</tr>
<tr>
<td>Repeat purchase rate of new customers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Five-week phone survey</td>
<td>11%</td>
<td>5%**</td>
</tr>
<tr>
<td>One-year phone survey</td>
<td>65%</td>
<td>46%**</td>
</tr>
<tr>
<td>Perceived promotional characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monetary savings</td>
<td>2.53</td>
<td>2.07</td>
</tr>
<tr>
<td>Entertainment</td>
<td>1.41</td>
<td>1.63**</td>
</tr>
<tr>
<td>Innovativeness</td>
<td>1.62</td>
<td>1.88**</td>
</tr>
<tr>
<td>Effort</td>
<td>3.70</td>
<td>5.00**</td>
</tr>
</tbody>
</table>

Table 4
Experimental design of Field Experiment 2.

<table>
<thead>
<tr>
<th>Studio</th>
<th>1st week</th>
<th>2nd–3rd week</th>
<th>4th week</th>
<th>5th–6th week</th>
<th>7th week</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PWYW</td>
<td>No promotion</td>
<td>40% discount</td>
<td>No promotion</td>
<td>Free sampling</td>
</tr>
<tr>
<td>2</td>
<td>PWYW</td>
<td>No promotion</td>
<td>Free sampling</td>
<td>No promotion</td>
<td>40% discount</td>
</tr>
<tr>
<td>3</td>
<td>Free sampling</td>
<td>No promotion</td>
<td>PWYW</td>
<td>No promotion</td>
<td>40% discount</td>
</tr>
<tr>
<td>4</td>
<td>Free sampling</td>
<td>No promotion</td>
<td>40% discount</td>
<td>No promotion</td>
<td>PWYW</td>
</tr>
<tr>
<td>5</td>
<td>40% discount</td>
<td>No promotion</td>
<td>PWYW</td>
<td>No promotion</td>
<td>Free sampling</td>
</tr>
<tr>
<td>6</td>
<td>40% discount</td>
<td>No promotion</td>
<td>Free sampling</td>
<td>No promotion</td>
<td>PWYW</td>
</tr>
</tbody>
</table>

4.1.3. Perceived promotional characteristics of PWYW vs. sampling

Conducting an independent group t-test, we find that monetary benefit played a marginally (t = −1.91, p < 0.10) greater role for trial in the free sampling condition (see Table 3). Asking the participants in the one-year survey how much the prospective savings had affected their decision to participate, we found that prospective savings were 18% more influential for free samples than for PWYW.

Furthermore, we find significant differences between the evaluations of the promotions regarding their entertainment factor. As Table 3 indicates, both promotions achieve high scores for entertainment; however, PWYW is rated to be significantly (+13.5%, t = 3.46, p < 0.05) more entertaining than sampling.

PWYW is also perceived to be a newer, more innovative promotional instrument compared with sampling. The respondents of the phone survey rated PWYW as being significantly more innovative (13.8%, t = 2.18, p < 0.05) than did the respondents under the free sample condition. This result may also explain the different word-of-mouth (WOM) behaviors between the two treatments. Participants under the PWYW condition stated that they spoke to 2.7 people (on average) about the PWYW promotion and to 2.5 people about the product. In contrast, the free sampling participants spoke to 2.1 people about the free sampling promotion and to 2.0 people about the product. Both differences are significant (WOM promotion: t = −3.04, p < 0.05; WOM product: t = −2.29, p < 0.05).

Finally, participation in the PWYW promotion is associated with significantly more effort compared with free sampling (t = 7.78, p < 0.05). Although the participants had to leave their contact information in both treatments, the PWYW participants were also required to determine the price to pay for the razors. Asking the PWYW participants how difficult it was to set the price, they answered with an average score of 3.70 that it was not too difficult (1 = “very difficult” to 5 = “not difficult at all”). However, PWYW was associated with significantly more effort than sampling, which required no additional effort.

4.1.4. Performance: trial and repeat purchases

Surprisingly, the number of trials in the PWYW treatment is not lower than that in the free sampling treatment. The (equally) high trial number in PWYW may be due to the innovative and entertaining attributes of the promotion.
Interestingly, we find significant differences in repeat purchases of new customers. From all new customers who in the first survey stated that they had not had any experience with the Gillette Fusion razor, 5.1% of the free sampling participants declared five weeks later (five-week survey) that they had already bought the corresponding razor blades, whereas the new customer rate was 10.8% under the PWYW condition. A chi-square test reveals that this difference (+111.8%) is significant ($\chi^2 = 3.82$, $p < 0.05$), i.e., PWYW leads to a higher repeat purchase rate of new customers than does free sampling. The results from the one-year phone survey confirm our findings. One year later, the overall repeat purchase rate of new customers was 55%, indicating that 55% of the respondents who were first exposed to the Gillette Fusion razor in our field experiment bought the razor blades within one year. The difference between free sampling and PWYW is marginally significant ($\chi^2 = 3.59$, $p < 0.10$), with 46% of new customers from the sampling condition and 65% of new customers from the PWYW condition, indicating that PWYW may screen out non-prospects. Asking the participants how much they were generally interested in the razors (Table 1) supports the screening assumption because the participants from the PWYW condition expressed higher interest in razors in general compared with the participants from the sampling condition ($t = 2.12$, $p < 0.05$).

4.2. Field Experiment 2

4.2.1. Data collection and measures

The goal of our second field experiment was to validate the results of our first experiment with another type of product and, additionally, to test PWYW vs. price discounts. We conducted the second field experiment with professional photo portraits at six comparable photographic studios. Within three weeks, each studio either (1) asked its customers to pay what they wanted for one print of a portrait, (2) offered a 40% discount or (3) offered one print free. Additional prints were sold at the original price. We counterbalanced the order of promotions to control for order effects (see Table 4).

All promotions were advertised with flyers, posters, A-boards as well as in the regional newspaper. Between the two promotion weeks, i.e., before rotating, we planned two weeks without any promotion. This approach avoided any overlapping of advertising and the current promotion. The costs for conducting this experiment were equal across the promotion types because there was no need for additional personnel for either the PWYW or the free sampling treatment. Customers who were interested in a current promotion were required to call the studios to make an appointment for their portrait to be taken. Following the photo shoot, every participating customer was surveyed. We asked all of the questions that were used in our previous study.

4.2.1.1. Descriptive statistics. A total of 138 customers participated in our study, with a distribution of N = 39/84/15 in the PWYW/sampling/discount condition. On average, the participants were 36 years old, and the majority were female (77.5% vs. 22.5%). Across all customers, the studios attracted 33.3% existing and 66.7% new customers.

On average, the participants paid €16.12 under the PWYW condition, which corresponds to 37% of the regular prices. This is a much higher percentage compared with the first field experiment (11.76% of the regular prices on average), which demonstrates that the monetary savings here are much lower than in Field Experiment 1. In sum, promotional revenues were €628.50 for PWYW, €385 for price discounts and €60 for photo prints in the free sample condition. A total of 57.2% of all the participants decided to order additional prints after the promotion (N = 22/44/13, PWYW/sampling/discount), and 62% of repeat purchases were made by new customers.

4.2.2. Perceived promotional characteristics of PWYW vs. sampling and discounts

Participants who received the discount perceived the respective savings to be less influential in their decision to participate in the promotion compared with the participants in the free sampling conditions (+25%, $t = 2.24$, $p < 0.05$). However, an independent group t-test revealed that the difference between discounts and PWYW was not significant (see Table 5). This finding may also be because the relative (to the regular) prices paid at PWYW were much higher in Field Experiment 2 than in Field Experiment 1.

A comparison of PWYW with either sampling or price discounts revealed no differences among the promotion types regarding the evaluation of entertainment. All promotions were evaluated to be highly entertaining. This evaluation may differ from our expectations and the results of our first study because all participants spent a considerable amount of time in the studio with the photographer. The personal interaction with the photographer may have been included in their evaluation, diluting the differences in the observed promotion evaluations for the product.

The participants in the price discount condition rated the promotion to be less innovative ($t = 2.50$, $p < 0.05$) than the participants in the PWYW condition did. However, the difference in innovativeness was not significant between PWYW and free sampling. Both promotion types were evaluated as being equally (high) innovative. A reason for this rather surprising result may be that it is not common to give away free portrait prints. In contrast, in study 1, we used a product (razor) that is quite often sampled.

Similar to our experimental design in study 1, the effort related to the survey was the same for all the participants, independent of the promotion type. However, also in study 2, the participants were required to set a price to pay. Again, we asked the participants how difficult it was to determine the price. Table 5 shows that the mean value for effort is significantly different (PWYW vs. sampling: $t = 11.46$, $p < 0.05$; PWYW vs. discounts: $t = 4.79$, $p < 0.05$) from the values for sampling and price discounts.

4.2.3. Performance: trial and repeat purchases

Reviewing the benefits perceived, it is salient that free sampling stands out compared with PWYW and price discounts. This result is precisely reflected in the participation rate: free samples appear to be most preferred, followed by PWYW and then by price discounts. Sixty-one
percent of the participants registered to receive one free print of a portrait, 28% determined the PWYW price, and 11% took advantage of the discounted price. On average, the participants in the PWYW condition paid 37% of the regular price, implying a 63% discount, which is much higher than the discount given in the 40% discount condition. However, due to the higher participation rate, lower prices paid in PWYW are overcompensated such that revenues made with PWYW (€628.50) exceed revenues generated with discounts (€385).

Furthermore, to capture the effect of repeat purchases of new customers, we observed additional print purchases. We found that 32% of all new customers bought additional prints in the sampling condition compared to 33% in the PWYW condition and 60% in the discount condition. The difference in new customer purchases was marginally significant ($\chi^2 = 3.19, p < 0.10$) when comparing the purchases in the PWYW and price discount treatment. However, the difference in the repeat rate of new customers between sampling and PWYW was not significant. This result may be due to a specific characteristic of our study. Because customers had to call the photo studios to make an appointment to participate, we most likely automatically screened interested customers. The results from our survey support this interpretation, as we did not find any significant differences across all treatments when asking the participants how interested they were in the product category.

Overall, PWYW appears to be an attractive alternative to discounts and free samples because PWYW generates additional revenues similar to discounts but, on the other hand, reaches more consumers than price discounts, similar to sampling. In our case, higher discounts taken by the PWYW participants were overcompensated by the greater trial rate compared with price discounts, leading the photo studios to achieve highest revenues with PWYW with €628.50, followed by €385 with price discounts and €60 with free sampling.

5. Discussion

Our results from both field studies suggest that PWYW is a relevant alternative to common non-price and price promotions such as sampling and discounts. More precisely, PWYW appears to be an intermediate solution between sampling and discounts that may yield the highest revenues.

5.1. Implications for marketers

Similar to sampling, PWYW is applicable for new and existing products to induce trial and attract new customers. However, PWYW is preferable when the product is too difficult to sample (e.g., services such as haircuts, massages and buffets). Costs of sampling may be higher considering a random distribution of samples leading to waste, but samples do not necessarily require sales personnel, whereas PWYW requires at least one person handling the sales. Depending on the marketer's objective, one has to weigh the costs against projected increases in revenues and WOM. Marketers may generate additional revenues from PWYW prices compared with zero prices when products are sampled for free. Furthermore, our results from field study 1 show that PWYW may yield a higher repeat purchase rate of new customers and that sellers using PWYW benefit from higher WOM behavior. With PWYW, the seller may also signal high quality (when the seller trusts his product quality) because consumers may align their prices after consuming the product.

PWYW is an alternative to price discounts in cases where marketers wish to attract a higher number of consumers. In Field Experiment 2, significantly more people participated in PWYW. In sum, with PWYW, the photo studios gained a higher number of repeat purchases among new customers and higher revenues than with price discounts. Based on our findings from field study 1, we would also expect higher WOM behavior with PWYW than with discounts. However, costs may also be higher because PWYW prices may not compensate for the costs of the products. Furthermore, PWYW may also require more personnel for its implementation. Therefore, the application of PWYW may be reasonable either when capacities are unused or when the product is associated with low variable costs.

Because the results of the two field studies underscore category differences, it appears that product promotions should be designed in such a way that they are innovative and entertaining to attract hedonic target groups. However, such a design should also limit the utilitarian benefit (monetary savings) or increase the effort associated with the promotion to screen out non-prospects. An interesting combination of PWYW and price discounts could be promotions in which customers may self-select their discount level from a pre-defined list of prices (see also Regner & Barria, 2009) or candidate discount levels (e.g., 20%, 30%, 40%). By designing the list of candidate discount levels, the company is able to influence the magnitude of the screening effect while still offering an innovative price promotion tool (Spann, Häubi, Skiera, & Bernhardt, 2012). Another option may be to distribute PWYW coupons. In this way, the promotion remains innovative and entertaining, but it implies additional effort by requiring customers to redeem the coupon.

5.2. Limitations and future research

Our study is experimental but exploratory in nature. Several limitations provide avenues for future research. First, the novelty of PWYW may fade as more companies use it. Second, a larger number of product categories are required to systematically relate categorical and promotional characteristics to promotional effectiveness measures. While such an effort may be prohibitive in field studies, laboratory studies could shed more light on these relationships in the future. Third, we did not systematically analyze the effects of PWYW, free samples and discounts on consumers' reference prices. Such effects appear to be a fruitful area for future research. Fourth, we did not observe consumers who decided not to participate in a promotion; we thus cannot test factors that influenced the decision to participate vs. not to participate. Fifth, another interesting aspect would be to analyze the behavior of those consumers who were already using the product before attending the promotions and to compare their behavior with the behavior of new customers.

Finally, future research may test alternative mechanisms and compare their effectiveness to that of PWYW pricing or sampling. Rebates, especially "free-after-rebates", for example, are a related type of price promotion where consumers pay the full price but can obtain a (possibly 100%) rebate of the price after the purchase if they perform a specific action, such as filling out a form and mailing it to the retailer or manufacturer. The required effort of this action may be similar or even higher than with PWYW. However, the screening component is different. In PWYW, the effort may screen out customers from obtaining the product in the first place. With rebates, the effort screens out the customers who will obtain the discounts—i.e., only those willing to incur the effort to obtain the lower price— from the others who will pay the full price. A comparison of these different but related types of price promotions appears to be an interesting opportunity for future research.

Acknowledgments

We would like to thank the editor (Jacob Goldenberg), the previous editor (Marnik G. Dekimpe), the associate editor (Koen H. Pauwels) and the reviewers for their helpful comments and suggestions. We would like to express our gratitude to Carsten Kortmann, Frank Burkhardt, Hanna Senftleben and Jochen Arndt from Procter and Gamble as well as to the owners of the photo studios, Claudia Köngsen from PHOTO PRESSE, and Roxana Kulawik for their valuable support in our field experiments. We thank Thomas Otter and the participants of presentations of this paper at the London Business School and Universidad Autónoma de Madrid for valuable comments.
Appendix A

Survey questions during field experiment (experiment 1)

Do you have any purchase or usage experience with the Gillette Fusion razor?
Do you purchase razors in your household?
Do you use razors?
If you have checked “yes” to Question 2 and/or Question 3, please state the last three razors you have purchased or used.
Please estimate the regular price of the Gillette Fusion razor.
Please rank the attributes (price, quality, handling, design), with 1 being the most important.
Please state your age.

Survey questions in the five-week phone survey (experiment 1)
1) Have you already tried the Gillette Fusion?
2) Please evaluate the Gillette Fusion based on these criteria:
   - Price
   - Quality
   - Handling
   - Design
3) How likely are you to buy corresponding razor blades?
   - Will continue using Gillette Fusion
   - Will switch to my previous razor
   - Already bought corresponding razor blades
4) Please state your maximum WTP for Gillette Fusion razor blades.
5) Why did you switch to Gillette Fusion? Why not?
   - Price
   - Quality
   - Handling
   - Design
6) How often do you participate in such promotions?
7) How attractive did you find the promotion?
8) Why did you participate in this promotion? Please distribute 100 pts on product or promotion.
9) With how many people did you talk about the promotion or about the product?
   - Product
   - Promotion
10) Only PWYW: How much would you pay if you participated again in the same promotion?
11) In the following, we will read statements about the promotion and Gillette. Please indicate how much you agree or disagree.
    Gillette is a fair provider.
    Gillette is good to me.
    Because of this promotion, I like Gillette better.
    With this promotion Gillette differentiates from others.
    This promotion suits Gillette.
    Gillette cares about its customers with this promotion.
    Gillette wants to convince me to buy an expensive product.
    This promotion is fun.
    This promotion is interesting.
    I felt anxious because of the promotion.
    This promotion was foolish.
    This promotion was fair.
    I felt good after the promotion.
    I look for specials in the supermarket.

Survey questions in one-year survey (experiment 1)
1) Do you still use the Gillette Fusion?
2) Do you still use your previous razor? Which one?
3) How often did you buy corresponding razor blades of Gillette Fusion in the past 3–4 months?
4) Have you bought the razor blades?
5) With how many people did you talk about the product or the promotion?
6) Have you recommended the Gillette Fusion?
7) Did your acquaintances buy the razor due to the recommendation?
8) How often do you shave?
9) How much are you generally interested in razors?
10) What is your maximum WTP for razor blades (4 pieces)?
11) Why did you switch to Gillette Fusion?
12) Have you recently participated in similar promotions?
13) Please rank the innovativeness of our promotion.
14) Only PWYW: How much would you pay if you participated again in the same promotion?
15) Have you bought the razor blades?
16) Only PWYW: How difficult was it to determine the price?

Appendix B. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ijresmar.2014.03.005.

References


